# Using STPA to Inform Developmental Product Testing

by

DANIEL R. MONTES
MAJOR, UNITED STATES AIR FORCE

B.S. Astronautical Engineering, United States Air Force Academy, 2003
M.S. Aeronautical Engineering, Air Force Institute of Technology, 2005
M.S. Flight Test Engineering, Air Force Test Pilot School, 2009

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**
at the
**MASSACHUSETTS INSTITUTE OF TECHNOLOGY**

February 2016



Signature of Author_____
Department of Aeronautics and Astronautics
October 23, 2015

Certified by _____
Nancy G. Leveson, Ph.D., Professor
Department of Aeronautics and Astronautics
Thesis Committee Chair

Certified by _____
Joshua C. Poore, Ph.D., Senior Technical Staff
The Charles Stark Draper Laboratory
Thesis Advisor

Certified by _____
Leia A. Stirling, Ph.D., Assistant Professor
Department of Aeronautics and Astronautics
Thesis Committee Member

Accepted by _____
Paulo C. Lozano, Ph.D., Associate Professor
Department of Aeronautics and Astronautics
Graduate Committee Chair

## Disclaimer

The views expressed in this document are those of the author and do not reflect the official position or policies of the United States Air Force, Department of Defense, or Government.

*In memory of Dave "Cools" Cooley and Mark "Dash" Graziano,*
*who made the ultimate sacrifice in the pursuit of knowledge;*

*for my amazing friend Stanley Mulenga,*
*the Lion;*

*and to Melissa,*
*thanks for all the fish.*

# Using STPA to Inform Developmental Product Testing

by

**Daniel R. Montes**
Major, United States Air Force

Submitted to the Department of Aeronautics and Astronautics on
October 23, 2015 in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Aeronautics and Astronautics

## Abstract

Developmental product testing currently evaluates system safety the same way it evaluates system performance: it attempts to isolate individual components' behaviors to evaluate their reliability. However, today's systems are often irreducible because of their complexity, leaving current practices ineffective at identifying safety deficiencies. Evolving to a modern systems-based hazard analysis is important for product development. Products stand to benefit during the testing stage, before initial fielding. In test, designs meet operation for the first time, and use practices and organizational influences both contribute to the safety of the system. By evaluating safety as an emergent property, hazards that emerge because of the testing process itself can be mitigated, and hazards that exist because of the inherent system design and use philosophy can be identified and traced throughout development and fielding.

System-Theoretic Process Analysis (STPA), developed by Nancy Leveson at the Massachusetts Institute of Technology, is a modern hazard analysis technique that identifies unsafe scenarios in a system in order to generate requirements to eliminate or control those scenarios. It improves on traditional reductionist approaches that treat accident causation only as a linear chain of events or probabilistic occurrence of simultaneous component failures (including human error). While systems-based and complete, STPA could benefit from additional guidance, particularly in the identification of human contributions to accidents.

The present research begins by extending STPA to include more guidance for the controller analysis, including refinements to the process model, fundamental human-engineering considerations, and socio-organizational influences. Next, Leveson's organizational control structure example is updated to include a test stage that serves as an intermediary between design and field use. Model inclusion criteria are updated, and Explicit-Influence Maps are introduced as a tool to understand the organization and aid in hazard analysis. Finally, this research investigates the U.S. Air Force developmental testing enterprise and applies STPA to a product test. Results are compared to that of the test-safety planning and reporting techniques traditionally in use, and utility is assessed with a research survey administered to developmental test professionals.

**Thesis Supervisor:** Nancy G. Leveson
**Title:** Professor of Aeronautics and Astronautics and Engineering Systems

***Keywords:*** *STAMP, STPA, system safety, hazard analysis, product testing, test safety, problem reporting, safety certification*

# Acknowledgments

None of this would be possible without Professor Nancy Leveson, who accepted me into this graduate program knowing I had an expiration date. We met in her office three-and-a-half years ago, and she asked me if I wanted to make testing safer. Around that time I had just finished revising a policy for mission-control-room standards in my organization with two very good friends, Joe Browning and John Casey, and test safety was already prominently on my mind. Thank you Nancy for giving me the opportunity to explore this world, the guidance to understand the problem, and the freedom to go after it.

I owe my technical advisor, Dr. Josh Poore, an incredible debt of gratitude for his support. I could not have had this experience without Draper Laboratory as my technical sponsor and John Scudiere, Steve Kolitz, Brenan McCarragher, and Chris Yu for inviting me in and providing me everything I needed to fulfill the program. My interests in engineering psychology and human performance were satiated during my work with Emily Vincent, Kim Jackson, Pete Lewis, Troy Jones, Jeff Zinchuk, Kevin Duda, and Jana Schwartz. A big thank you to Dr. Jeff Jungemann, for sitting in my thesis defense and, through the years, encouraging me to remember the big picture in the Air Force. Thank you all and especially Josh for the wealth of scientific knowledge and exposure to state-of-the-art applications shared with me at Draper.

An immense thank you to committee member Professor Leia Stirling, who always asked the right questions for me to consider the way I presented my research results. I was privileged also to receive guidance and partake in engineering academics taught by amazing leaders in their fields including Larry Young, Dava Newman, David Mindell, Charles Oman, Julie Shah, John Flach, John Carroll, Richard de Neufville, John Hansman, Kerri Cahoy, Missy Cummings, and Divya Chandra. My appreciation also goes out to Professor Sheila Widnall who helped evaluate my research proposal, and to Professor Oliver de Weck for teaching me the systems-engineering process and allowing me to participate in a superb design-and-build engineering competition my first year at MIT.

Our diverse student lab at MIT embodies the wide-ranging industries that STAMP has impacted. I am grateful to John Thomas, Blandine Antoine, John Helfrich, and Cody Fleming for nurturing us junior members through, William Young for teaching me how to navigate this type of endeavor as a military officer, and the rest of the crew: Adam, Andre, Aubrey, Blake, Cameron, Connor, Dajiang, Emel, Jonas B., Jonas H., John S., Kip, Meaghan, Seth, Soshi, and Yonatan. To all the new folks getting to the lab, you have a great group of people there to welcome you. Furthermore, Sophia Hansefus, Julie Finn, Beth Marois, and Marie Stuppard are the lifelines of this program and one of the biggest reasons any of us survive this ordeal, so much appreciation goes to them for making everything work. A big thanks also to Thelma for providing me the late day coffees and Craig for the colorful late night conversations when he would take away my empty coffee cups!

The student life at MIT was filled with energy and clarity through fellow travelers. Whether we were schlepping through together during classes, group projects, student council, movie nights, restaurant weeks, jam sessions, races, ski trips, the rock gym, or the pub, the journey was always illuminated by folks like Aaron, Abhi, Alexander, Allie, Ana, Andrew, Annie, Bassel, Becky, Bo, Bobby, Brad, Brandon, Celina, Charles, Clifton, Dilip, Eddie, Fernando, Forrest, Giuseppe, Gwen, Holly, Ioana, Irene, Jack, Jared, Lawrence, Louis B. Louis P., Luke, Margaret, Marc, Matt, Narek, Nikhil, Nora, Oli, Pat, Pearle, Pem, Pete, Phil, Pratik, Raquel, Remi, Rich, Sathya, Seb, Sherrie, Shervin, Steve A., Steve O., Sydney, Tim, Tony P., Tony T., Torin, Vishnu, Whitney, and Yango. Additionally, it was a pleasure and honor to see the younger generation of military officers start their careers here, and I enjoyed interacting with all the Lieutenants including Bryan, Caleb, Casey, Dan, David, Dustin, Dylan, Evan, Greg, Jack, John, Kevin R., Kevin S., Mark, Meghan, and Travis. Special thanks to Steve Fino for making that opportunity possible and setting the example. To all these fine folks, I cannot be grateful enough for their friendship and all the kind things they have done for my family, including walking our dogs countless times and even throwing us a real wedding reception!

I think of my friend Nick Chung anytime I reflect on the way this research matured in scope. He was a not only a lab mate at MIT but a fellow engineer during our former lives as testers. He and I spent many whiteboard sessions mapping out our interpretations of the various research gaps we sensed in our similar areas of interest. I truly appreciate the inspiration he gave me, along with so many colleagues in the military who were a continual source of support and feedback. Nick "Hammer" Helms, John Sciuto, Andy Bogusky, Tim "Astro" Cullen, Matthew Domsalla, Jess Buchta, Pierre Romeo, Paul Robinson, Patricia Dunavold, Kelly Wolfe, Stephanie Storch, Erik Nelson, Mike Berard, Mike Kinney, Jim McCorduck, Tomasz Stec, and Scott Jones gave me a great perspective of the problems facing unmanned vehicle systems. Paul Waters, Doyle Janzen, Rob Warner, Kevin Wodarck, Andrew Hansen, Bryan "Groucho" Duke, Tom "Sulu" Hill, "Evil" Bill Gray, Brian Donnelly, Matt Clark, and Kerianne Gross opened huge doors and shared a wealth of research and test experience. Jason Carter, Dan "Animal" Javorsek, Jason Bartolomei, Luke Cropsey, Sean Estrada, Harris Hall, and Aaron Tucker gave me valuable advice for preparing school applications. Theresa Dearth helped me and many others make it into the schools we pursued, and Luke Whitney, William Friedrich, Carl Corvin, Dan Clepper, and Mike Rosenof gave us fantastic personnel support. Clint "Void" Armani, Steve "Burns" Ross, Jose "Hummer" Gutierrez, and Sean "Woody" Musil provided mentorship and encouragement from Test Pilot School. Thank you all for everything.

A venture like this would not be possible without the full support of leadership. My sincerest gratitude to Dr. Joseph Nichols for sitting in my thesis defense as great ear for the test world and for offering his insight. Neal Barlow, my undergraduate thermodynamics professor, supported me all the way to admission to this fantastic graduate program. Paul King and Mark Gruber, my master's program advisors, encouraged me to pursue a doctorate; Chuck Wolfe, my first commander and a shining example of real leadership, took me into his office my first week on the job and learned everything he could about me; he knew how to fit the work to the person, and he bent over backwards to take care of his troops. Vince Parisi and Steve Fernandez gave me the freedom to run my own research and development team a lifetime ago, a type of flexibility not often afforded to young officers. Ryan "Rooster" Osteroos is the reason I became a tester and a continual source of encouragement and inspiration. Noel "Shamu" Zamot modernized Test Pilot School to accommodate the next generation of defense technologies and evaluation methodologies; he has always had the big picture in mind regardless of short-term politics, and he fought to make my academic opportunity at MIT possible to begin with. Doug Wada is another big-picture thinker who backed me up anytime the management could not see the forest from the trees. Thanks always to Jim West, my skydiving mentor, and Tracy Scanlan, my flight-test mentor, for blazing the trails so many have followed.

Although I cannot name the survey participants that provided the data in this thesis, I must express enormous appreciation for the time they devoted to expressing their honest thoughts on test safety. Meaningful change is not possible without obtaining feedback from front-line practitioners, but there will always be lower offices in any large organization that are not capable of being receptive to this information. I am thankful we have the opportunity now to treat safety with the open mindedness it deserves.

Thanks Marc for joining me at the border every time I needed to clear my head. Thank you Stanley for reminding me every day what real courage and determination is. Thanks Keisha for being an amazing inspiration for anyone wanting to make a difference outside their comfort zone. Thank you Mom, Dad, Dave, and Darren for always being supportive. And last, not least, a big dolphin hug to my wonderful wife Melissa for the love, wisdom, curiosity, and beauty you bring into our lives every day. Thanks for feeding the dogs all those times I was absorbed in the writing. I love you all.

# Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| 412TW | 412th Test Wing |
| ADS-B | Automatic Dependent Surveillance-Broadcast |
| AF | Air Force |
| AFB | Air Force Base |
| AGL | Above Ground Level |
| AFI | Air Force Instruction |
| AFMC | Air Force Materiel Command |
| AFPD | Air Force Policy Directive |
| AFRL | Air Force Research Laboratory |
| AFSEC | Air Force Safety Center |
| AFTC | Air Force Test Center |
| AIB | Accident Investigation Board |
| AIP | Aircraft Information Program |
| ATC | Air Traffic Control |
| CA | Control Action |
| CAST | Causal Analysis using Systems Theory |
| CC | Communication |
| CRM | Crew Resource Management |
| CSE | Cognitive Systems Engineering |
| CTA | Corrective Action |
| DOD | Department of Defense |
| DODD | Department of Defense Directive |
| DODI | Department of Defense Instruction |
| DR | Deficiency Report(ing) |
| DT | Developmental Test (and Evaluation) |
| $E_C$ | Expected Casualties |
| EIM | Explicit-Influence Map |
| FAA | Federal Aviation Administration |
| FB | Feedback |

| | |
|---|---|
| FCIF | Flight-Crew Information File |
| FHA | Functional Hazard Analysis |
| FL | Flight Level |
| FMEA | Failure Modes and Effects Analysis |
| FTA | Fault Tree Analysis |
| FY | Fiscal Year |
| GCS | Ground Control Station |
| GMP | General Minimizing Procedure |
| HF | Human Factors |
| HFACS | Human Factors Analysis and Classification System |
| HSI | Human-Systems Integration |
| IBE | Item(s) Being Evaluated |
| IM | Indirect Measure |
| ITP | In-Trail Procedure |
| MABA | Men Are Better At / Machines Are Better At |
| MAJCOM | Major Command |
| MFOQA | Military Flight Operations Quality Assurance |
| MIT | Massachusetts Institute of Technology |
| MP | Minimizing Procedure |
| MV | Manned Vehicle |
| NAS | National Airspace |
| NASA | National Aeronautics and Space Administration |
| O&M | Operations and Maintenance |
| OODA | Observe, Orient, Decide, Act |
| ORM | Operational Risk Management |
| OSH(A) | Occupational Safety and Health (Administration) |
| $P_C$ | Probability of Casualty |
| PDF | Portable Document Format |
| PHA | Preliminary Hazard Analysis |
| PRA | Probabilistic Risk Assessment |
| QA | Quality Assurance |
| R&M | Reliability and Maintenance |

| | |
|---|---|
| RA | Recovery Action |
| RCC | Range Commander's Council |
| ROA | Remotely Operated Aircraft |
| ROV | Remotely Operated Vehicle |
| RPA | Remotely Piloted Aircraft |
| RPV | Remotely Piloted Vehicle |
| SA | Situation Awareness |
| SIB | Safety Investigation Board |
| SME | Subject Matter Expert |
| SSSI | Single Sensor, Single Instrument |
| STAMP | System-Theoretic Accident Model and Processes |
| STPA | System-Theoretic Process Analysis |
| STPA-RC | System-Theoretic Process Analysis, Refined Controller-Analysis |
| SDT | System During Test |
| SUT | System Under Test |
| THA | Test-Hazard Analysis |
| TPS | Test Pilot School |
| TRL | Technology Readiness Level |
| TSM | Time Safety Margin |
| TSPI | Time, Space, and Position Information |
| TTPs | Tactics, Techniques, and Procedures |
| UAS | Unmanned Aircraft System |
| UAV | Unmanned (Unpiloted) Aerial Vehicle |
| UCA | Unsafe Control Action |
| UUV | Unmanned Underwater Vehicle |
| UV | Unmanned Vehicle |
| V&V | Verification and Validation |
| VMC | Visual Meteorological Conditions |
| VUE | Visual Understanding Environment |

# Chapter 1

# Introduction

> "There are no accidents and no fatal flaws in the machines; there are only pilots with the wrong stuff ... no single factor ever killed a pilot; there was always a chain of mistakes."[1]
>
> —*Tom Wolfe*

Traditionally, safety experts assume that accidents are caused by component failures. This view was popularized in the early-to-mid twentieth century when machines were composed of electrical and mechanical parts connected with simple relationships; having more reliable components usually resulted in a measurable decrease in loss of life and increase in mission effectiveness. Human operators were originally expected to be absolute experts in the machinery, tools, and devices they used and to operate their system with perfect skill. If a component failed or degraded and the operator could have caught or mitigated the problem, it was the human who was blamed for letting the situation become a loss.

In the traditional view, accidents are caused by a linear chain of events or simultaneous failures; component failures and human errors are assumed to be probabilistic with a fixed rate. Whether or not this model of accident causality was warranted in the past, the complexity of modern systems requires taking into consideration the interactions between components and emergent system behavior. It also demands reconciling the human's role into that of being part of the complex system, and not simply an operator with the "right stuff" overseeing machines with predictable behavior. In today's highly interactive systems there may in fact be fatal flaws in both the design and operating procedures. With a more modern approach, safety considerations can be engineered into systems at design and evaluated during testing. Assumptions and uncertainties about both the design and the operating procedures can be traced throughout the transition between testing and field use. Hazards can be identified as they pertain to the systems being tested as well as the testing activities themselves, and they can be mitigated as stakeholders deem appropriate.

---

[1] *The Right Stuff* [1, pp. 25–26].

## 1.1 Research Purpose

The purpose of this research is to:

> *Improve the ability to assess system safety during developmental product testing and standardize the applicability of hazard findings between the design and field use of the product.*

## 1.2 Research Background

The traditional approach to hazard analysis views safety as a function of component reliabilities and reduces the narrative of an accident to a linear chain of events, where each event occurs with the probability of some failure or error. This encourages the leveraging of available past experience, practitioner wisdom, and conventional reliability engineering to forecast risk. While this approach is not wholly ineffective as a method for accounting for accidents, it is inconsistent, inefficient, and incomplete. A linear chain-of-events model reflects what Taleb calls the *narrative fallacy*, which drives people to draw an "arrow of relationship" onto a sequence of events [2, p. 64]. People often retrieve stories of similar events from the past to suggest mitigations for the future, but this creates an opportunity for reductionism and subjectivism. *Hindsight bias* is particularly endemic in traditional approaches, where even experts find critical points in a narrative of past events in which to inject common sense on what components of a system should have been behaving appropriately or what decisions an operator should have been making [3]–[5]. When simple component relationships do exist—as they did early in human history before elaborate sociotechnical systems[2] existed—breaking the chain of component failures is enough to stop the "arrow".

While narratives can be stretched to fit new situations, they do not add more information for understanding those situations. Additionally, humans are frequently blamed for accidents because they are less understood than the other components of systems— they are unpredictable. Yet, they are seen as providing a last line of defense, hence the ultimate level of accountability. The traditional safety view cannot account for dynamic component interactions and non-linear behavior in elaborate, software-intensive systems. Mitigations from past experience implemented to halt simple failures based on a linear accident model can instead introduce new hazards. A more appropriate accident model treats safety as a control problem, both over the design and the use of a system. This modern view is capable of treating a human operator as part of the system itself and thus a contributing component—a decision-making controller—instead of an overseer that might allow the entire system to fail at some estimated human-error rate.

System-Theoretic Accident Model and Processes (STAMP) provides an approach for treating safety as a control problem. It was developed at the Massachusetts Institute of Technology (MIT) by Nancy Leveson as a new accident model; it is different than the traditional model of failure chains and component reliabilities (including reliabilities of

[2] This term was coined in the 1950s to refer to complex work domains influenced by both technology and by human behavior and social infrastructures [6].

people and software) [5]. STAMP treats safety as a property of the whole system, not any individual component. The prevention of undesirable losses is incorporated as a top-level set of system requirements, and with them designers and practitioners can generate appropriate constraints within the functional system behavior. Requirements and constraints can be managed throughout the development and use of the system using a systems-engineering process. By treating safety this way, it can be dynamically managed by everyone throughout the life of the system.

Where the traditional model is implemented through techniques like Failure Modes and Effects Analyses (FMEA) and Fault Tree Analyses (FTA), STAMP is implemented through its own engineering techniques used for various purposes such as accident investigations or concept development. One technique, which is the focus of this thesis, is called System-Theoretic Process Analysis (STPA). STPA is a hazard-analysis that identifies hazardous scenarios from a system design in order to generate functional system requirements to eliminate or control those scenarios. Engineers can use the system model to identify hazardous behavior without depending on past incidents to inform the analysis.

The evolution of STAMP was influenced in part by risk management as modeled by Jens Rasmussen, who took a cognitive-science perspective that viewed risk as an operational control problem [7]. In the field-use of a product or system, he represented the safety-control structure as including all levels of society and policy down through the organization performing the operations. He also sought a convergence of human sciences, decision theory, and management research into the discipline of cognitive systems engineering [8]. Safety practice could be accomplished by a systems approach that gives stakeholders a functional view of their organization and operations, allowing behavior constraints and work boundaries to be implemented and enforced.

Leveson improved Rasmussen's model by treating safety as a combined problem of development and field-use. In formulating STAMP, she presented a generalized systems-engineering approach in which stakeholder and designer requirements and constraints are communicated between these two major stages of a product lifecycle, with a functional control structure representing them both. She developed *intent specifications* as a method for designers to support both developmental and operational problem-solving and software evolution based on systems theory and cognitive psychology [9]. This approach implements Rasmussen's philosophy that the designers and operators cooperate in decision-making, with the designers communicating intent to the operators, and the human operators completing the design [10].

Organizations that view safety as a traceable metric throughout the life of a system sometimes employ system-safety guidelines like those in MIL-STD-882 [11]. A system-safety management program does at least ensure that the safety requirements at each stage are being adhered to, with some information flow between stages. They might apply particular risk mitigation practices at various segments within each product stage, especially if they must meet government safety regulations. Depending on the rigor of the product organization's systems-engineering process, the amount of communication among practitioners of these stages may vary. The design stage typically includes component reliability analyses and failure-chain scenario cataloging. The field-use stage applies many

risk-mitigation initiatives for the human operator(s), based on the old view that portrays the operators as overseers with an estimated error rate.

In many industries, developmental testing is a lifecycle stage in its own right that exists between design and field use. The product-testing setting—where design first meets operation—presents an area that would benefit from a study of its safety underpinnings. Safety in this stage is an explicit practice with two goals. The first is to verify the safety of the system as designed and intended for field use. This includes confirming assumptions that went into design modeling and intended operating procedures. Airworthiness certifications for airplanes would be an example of this. The second goal, known as *test safety*, is to ensure the safety of the test process itself, including techniques, configurations, and approaches for discovering inaccurate design assumptions. In a test setting, sometimes models and reliability estimates are all that are available on which to plan.

The test enterprise often depends on a very small number of past accidents from similar systems to provide enough information to forecast the future—even when technology and applications are constantly evolving. Without a modern view of systems and safety, designers, planners, operators, and maintainers apply their own safety practices (usually reliability-based) independently from one another. The system implementations based on post-hoc narratives are carried forward by requirements managers, design engineers, and testers into the field. However, when new technologies are tested for the first time, there may not be experience to guide preparations. Forecasting hazard scenarios becomes difficult, designing their mitigations even more so. A shift to a systems-based safety practice can provide an improved framework for documenting hazards in the absence of historical data.

STPA offers an improved technique for system-minded safety planning. It incorporates—but does not depend on—past experience to aid in identifying potential hazardous behavior. It can provide a framework to model the design of a system with the operations of the system, whether it be in testing or in the field. That framework can serve as a common planning and communication tool throughout all the stages of product life. With common planning tools come faster avenues of organizational feedback to stakeholders during development and end use.

Research organizations are advancing autonomy research toward system concepts in which software is poised to make higher-level decisions including self-organizing behavior and even value-based choices. Complex, emergent system behavior is already outpacing the capabilities of traditional safety-planning methods, which analyze the physical components of a system and make attempts to predict mission degradations. A functional, top-down systems-based analysis allows development organizations to control for emergent behavior and treat safety as a system property that can be managed across product-development stages as a common goal.

### 1.2.1 Gaps

#### *1.2.1.1 The STPA Human Controller*

The following challenges apply to STAMP/STPA in general:

- The *process model*, one of the main concepts of STAMP, does not capture types of system abstraction that human controllers require to contribute adaptivity to system behavior.
- Fundamental human considerations are not explicitly considered in the controller analysis (e.g., workspace factors, variability of personal traits).
- There is no current method to model the impact of social and organizational influences on the controllers within the operating process.

STPA forces analysts to consider functional behavior in the context of well-defined constraints. It also goes beyond the tendency to simply state that a human operator failed by identifying discrepancies in the human's process model (also called *mental model*) and flawed decision-making. STPA, although advanced in terms of safety analysis, still oversimplifies the human's role in complex systems because it is currently posed similarly to analyzing a machine controller's model and decision algorithm.

Humans contribute to the adaptivity of systems, which demands a more refined analysis of potential hazardous decisions. Human mental models contain more types of information about the system than a machine's and develop using more sources of feedback. Human performance and decision-making are also influenced by additional factors not present for machine controllers. These include human-performance considerations as well as socio-organizational factors that should be included in the analysis.

#### *1.2.1.2 Safety Management in Developmental Test*

The following challenges apply to a product's developmental testing stage:

- There is inconsistent expert knowledge at any given test-safety review board.
- There may be minimal expertise in new technologies (e.g., software, autonomy).
- The test-safety planning process does not use common visual aids in its documentation.
- Test engineers do not have a consistent method of tracing undesirable behavior or potential design flaws to effects on the system within the context of field use; this especially affects human-engineering experts, who cannot ignore the relationship between operating philosophy and system design.
- Problem reports tend to be reductionist (e.g., manufacturing error, component defects) and do not consistently explain system impacts through anything but written narratives.

- STPA control structures do not acknowledge an explicit product testing stage to capture the particular sociotechnical dynamics found in a test enterprise.

When products undergo formal testing, a modern approach to safety planning would help testing to be safer and provide more useful safety information about the product. Testers must first and foremost ensure that local testing activities are safe before assessing the inherent safety of the product. Depending on the size of the organization, the group that reviews a product's test-hazard analysis might be different from case to case. This results in an inconsistency of expertise at review boards. If the planning also leans heavily on subject-matter experience and wisdom, no two safety review boards will reach the same conclusions about how to best proceed with hazard mitigation for the testing. Furthermore, newer technologies, such as software and autonomy, may take quite some time before a pool of experts is even created.

Although design artifacts such as vehicle schematics are often provided as a reference to substantiate a product's hazard analysis, a further step would be to construct a common visual aid to be used by all test planners, regardless of background or experience. A shared model of the system during test allows a common set of systems-based principles to remain at the center of any discussion or review, and it enables different subsets of experts to collaborate on the common task (i.e., safety planning) [12].

There are often no common safety protocols that cover the entire development life of a product. Test-safety planning is often a function internal to a test enterprise, and its outputs might not always be considered part of the more general system-safety documentation. Currently, test engineers attempt to understand a system's design and functions based on a limited time to learn. As such, test engineers can only predict the use implications of the design and its behavior based on their limited experience, or if time permits, by speaking to field-use representatives. Discipline-specific engineers—particularly human-engineering experts—do not have a consistent method for tracing undesirable behaviors or potential design flaws that affect the system as a whole. Further handicapping test engineers is the fact that procedures for appropriate use in the field, and safe operating techniques and restrictions are validated and implemented only after extensive field use. Additionally, the enterprises that use the product must follow several independent regulatory and/or company policies regarding various aspects of safety (e.g., workplace/occupational, design/certification, operational risk management), which can further discourage a common safety protocol.

There is no reason that applying a modern, systems view of safety during testing should begin or end during the test stage. A process should be in place for practitioners from all stages of development to share a system model and/or set of specifications for putting safety findings within any product-development stage into perspective. Developmental test is often the first stage of product development in which both design and use techniques of a new system are exercised; furthermore, there are often unique features during test such as instrumentation and specialized software that may be different from the production system. A common framework for hazard analysis can not only improve the consistency and rigor of local test-hazard planning but also assure system

safety from design to use philosophies if properly managed—with safety practitioners contributing to maintaining its validity during all stages of development.

At the onset of testing, the majority of test-safety planning focuses on documenting the safety of the test activities. Following that, testers might be able to identify system-design and component-defect issues. Ultimately, the power to inform the safety of the product relies on the quality of problem reporting; testers try to prevent undesirable features of the system from making it to the field before field-use operators find the same mistakes. For government procurements, a type of problem reporting called deficiency reporting (DR) is mandated by federal regulations for quality assurance (QA) [13]. At their core (and historically), DRs are meant to catalog design and production defects of physical components. DRs have increasingly been used to attempt to capture more systemic issues in products, but their structure does not offer engineers the power to link findings to the impact on system function through anything but written narratives. The lack of traceability to design makes it difficult to develop remediation solutions.

STAMP is a powerful accident model that can help system managers build frameworks that treat safety as a top-level system property, traceable throughout all stages of product life. It offers the modern system-safety view that can address the research challenges discussed above. Currently, the most generalized example of a STAMP organizational control structure, produced by Leveson shows interactions between the design and field-use stages as low-fidelity communication channels containing product maintenance and evolution information [5]. The existing generalization would benefit from having a dedicated test stage between product design and product field use. By acknowledging a test stage, information about design and use assumptions can be appropriately highlighted and maintained between stakeholders within the different enterprises that contribute to the product development and field use.

### 1.2.2 Objectives

This research has two complementary objectives:

1. Extend STPA to better examine human controllers in the hazard analysis.
2. Provide a common framework for test-safety planning that addresses both the safety of the test process and inherent system safety.

The first objective addresses the gaps in Section 1.2.1.1 and aims to update the STPA controller analysis to include more refined system information in the mental model, add fundamental human-engineering considerations, and present a method to identify socio-organizational influences to the operating process. This objective addresses and is most applicable to human controllers. However, the additional guidance that is developed for the controller is generalized to analyze any intelligent controller in complex work domains. As autonomous controllers become more capable, their process models will include system information that human mental models already currently incorporate. This STPA extension improves the methodology of the hazard analysis by increasing the rigor in which causal scenarios involving both humans and autonomous controllers are identified.

The second objective addresses the gaps in Section 1.2.1.2 and aims to apply STPA to build a systems view of testing. STPA inherently provides a visual aid (in the form of a safety-control structure) that is shared by experts from various disciplines for the common purpose of safety-hazard identification and mitigation. This can improve local test-safety planning and ensure safety of the test process, especially for systems that incorporate new and emerging technologies. Additionally, this objective produces an updated generalized example of a STAMP organizational control structure so that system safety (including certification requirements such as airworthiness) is also assumed within an explicit test stage. If a system's control structure is created early in concept and design, it can benefit all stages of development, including test and field use. Operators and discipline engineers at each stage are able to trace their concerns to the top level system requirements and constraints. Testers are able to discuss deficiencies in a more consistent manner that is communicated in the context of field use.

## 1.3 Research Application

STPA is a powerful hazard analysis technique that has been successful in many industry domains, including aerospace, medical, and defense [5]. The U.S. Air Force (AF) was chosen for the application of this research. The AF contains an established organization for product development and acquisitions called the AF Material Command (AFMC). AFMC works officially with field-use organizations in the AF to steer the cradle-to-grave lifecycles of the military systems in use by the AF. Because it is a government entity, there are regulatory requirements for safety throughout the various stages of product development and use. There is also a dedicated product-testing enterprise within AFMC called the AF Test Center (AFTC). AFTC was created in 2012 as part of a reorganization that formalized a unique professional enterprise to shepherd each major stage of a system lifecycle. This organizational mindset reinforces the concept that a test stage is an explicit part of system life, driven by unique constraints and executed by expert test professionals.

AFMC could benefit from a modern system-safety approach as newer products are pushing the edge of technological possibility in the realms of software and autonomy. Systems like unmanned vehicles (UV) and flying drones have become apparent to the public eye due in large part to the large push by the AF over the last two decades—followed by other military services—to incorporate autonomy-capable platforms into their inventories. Autonomy-capable systems bring more complexity to the already elaborate aerospace domain, and traditional safety approaches are not enough to mitigate hazards that emerge from complex behavior. Therefore, UVs provide a good case for research application.

This research is generalizable to any industry wishing to acknowledge an explicit test stage, and the methods demonstrated here can be reproduced by subject matter experts (SME) within their respective industries.

### 1.3.1 Autonomy and Flight

Inevitable cultural resistance to change introduces challenges to technology adoption. For example, typewriters—a technology to replace handwriting and the premier example of automation in their time—were developed with the QWERTY keyboard layout in the 1870s. The layout was developed based on English word-letter relationships to lower the probability of typists hitting two adjacent keys simultaneously, as this caused sticking problems in early typewriters. The problem was mechanically fixed shortly after the invention of the typewriter, and modern electronic devices have all but replaced typewriters in most applications. However QWERTY remains to this day, despite more efficient keyboard layouts having been proposed [14]. Acceptance of improved technology occurs, albeit with resistance, when the value is great enough. The aerospace industry has slowly incorporated new technological trends as users begin to depend on the capabilities of the machines.

The Air Force Research Laboratory (AFRL) is the technology development enterprise under AFMC. The AFRL vision for developing autonomy technology is "Intelligent machines seamlessly integrated with humans maximizing mission performance in complex and contested environments" [15, p. 4]. Objectives include advancements in human-machine teaming, shared perception, self-governing teams of machines, robust communication, and flexible decision-making paradigms. While useful definitions for automation and autonomy are discussed further in the next chapter, the important emphasis here is that more and more functions once reserved for humans are meant to be performed by machine processes as technology improves. These functions are beginning to include value-based decisions, not just performance optimization. Products incorporating these advanced technologies and concepts are a critical application for system-theoretic safety models like STAMP.

An evolving application of autonomy is the UV. This is, most simply put, a vehicle that operates without an onboard human controller. The Dallas-Fort Worth Airport's Skylink tram—an "automated people mover"[3] that takes passengers between flight terminals—is an example. Depending on the work domain, UVs might be acknowledged by many variant terms including unmanned/unpiloted aerial vehicle (UAV), unmanned aircraft system (UAS), unmanned underwater vehicle (UUV), remotely piloted vehicle/aircraft (RPV/RPA), remotely operated vehicle/aircraft (ROV/ROA), or simply "drones" [16], [17]. UVs might be continuously controlled via radio or cable transmissions, allowed to operate independently of human supervision, or some mode in between. UVs pose interesting testing and employment challenges due in part to the guidance, control, and autonomy software algorithms they execute as well as the complex sensor and payload configurations made possible by the additional design space formerly reserved for on-board human life support. There is also a difference in environmental awareness between the human operators of traditional manned vehicles (MVs) and UVs. The pilots of MVs typically occupy cockpits and receive visual, aural, tactile, and vestibular feedback from the vehicle and its domain, whereas a UV pilot/supervisor is located in a separate place,

[3] Bombardier, Inc. provides the technology.

away from the operating domain and sometimes experiencing communication delays or discrepancies.

The employment and traffic-management of air vehicles is one of the most complex sociotechnical systems in existence. One challenge affecting the aviation community is the safe control of shared airspace between flying UVs and MVs. In the military, shared airspace deconfliction focuses on two areas. One concerns larger scale, runway-employing UVs. These vehicles operate in high density with MVs in the airfield traffic pattern and departure/arrival flows, where MVs and traffic controllers traditionally use visual deconfliction techniques. Once up and away, UVs are able to deconflict using standard traffic control and instrument procedures. The in-transit airspace is less dense than an airfield pattern, and sometimes the UVs are also capable of attaining very high altitudes that aid in thinning out the airspace. The other area concerns smaller scale UVs that operate at low altitudes (below roughly 3,500 feet), typically serving ground support tactical functions. Although they take off and land in remote, austere areas in low proximity to other flying vehicles, once up and away they operate at a high density with MVs (usually helicopters) as well as each other. Physical avoidance again traditionally relies on visual deconfliction techniques. Midair collisions occur with some frequency, usually resulting in the loss or damage of one or more UVs and very rarely an MV [18], [19].

The Federal Aviation Administration (FAA) is also making efforts to integrate the civil national airspace (NAS) for growing commercial and private use of UVs. This goal, along with the anticipated future increase of in-transit traffic density, will require a deconfliction strategy covering various altitudes and flight regimes [20]. Commercial UVs will utilize more than just the high altitudes of large military surveillance drones and low altitudes of small support military UVs. Thus, the focus for shared NAS deconfliction will not be limited to airfield patterns and low level regimes. As military UV employment continues to increase, military use of airspace will also need to account for increased densities, requiring solutions similar to the FAA.

A second challenge is the assurance that UVs will not harm ground assets (human injury/death or material damage). In the military, this concern is typically not as pronounced during field use—abroad and/or during combat or emergency operations—as it is during stateside training and testing, where government employees, government assets, and nearby civilian populations exist stateside under the flight area where new air vehicle designs might be unproven. With the FAA, this concern for assuring safety of people and things on the ground exists equally in the testing, training, and field-use stages. The operational national airspace encompasses populated as well as unpopulated areas.

A recent investigation by the *Washington Post* tallied 418 UV Class A and B mishaps[4] by U.S. military systems since 2001. Out of roughly four million hours of military UV flight, this equates to an accident rate of approximately 10.5 major mishaps per 100,000 flight hours.[5] About one half of the mishaps happened in a major theater of operations (Afghanistan and Iraq), one quarter in a minor theater overseas, and the remaining one

[4] See Appendix A for mishap class definitions.
[5] Data were obtained from public record. See Appendix A.

quarter in the continental United States. The majority of the military mishaps were with the Army, where UVs sustained an accident rate ten times that of their MV fleet in the last year alone [19].

In the Air Force (AF), the five-year Class A mishap rate for UVs during fiscal years (FY) 2009–13 was approximately 4.3 mishaps per 100,000 hours, more than double the rate seen in F-15 and F-16 fighter planes[6] (1.8 mishaps per 100,000 hours) during the same period. The common UV mishap factors cited include human error, mechanical defects, unreliable communication links, and a limited ability of the pilot to detect collisions or undesirable positions/attitudes.

While causal factors cited in UV accident reports are discussed further in the next chapter, the tendency to assign blame to human pilots is worth mentioning because it is persistent and not new. In aviation over the last 60 years, 70 to 80 percent of accidents have been attributed to human error [4], [5], [21]. A recent aviation safety study recommended a focus on operator fatigue prevention, crew communication techniques, increased training to understand automation, and traceable safety management policies in order to deal directly with human error [22, pp. 34–38].

Improvements targeted at improving UV safety are applied with varying degrees of practicality through three domains: policy (e.g., airspace partitioning, right of way rules, safety rules), autonomy algorithms (e.g., long range path optimization, short range sense and react), and human engineering (e.g., controls/displays, decision-making tools, training, team management). Each of these disciplines provides focused solutions toward large complex challenges. Because there is an interdependence among improvements, formal methods to model the larger system and account for dynamic contributions of combined disciplines would go further to help identify emergent properties of the whole design. A STAMP approach to safety is able to model the entire sociotechnical enterprise—not just the individual pieces (like a single aircraft)—and provide causal scenarios for hazards that do not blame human error or component failures.

Advancements in autonomy and the presence of multiple capable human and artificial controllers in complex systems necessitate rigorous methods to classify information within the work domain. In order for designers to build desirable *robust* and *flexible*[7] capability into a system, they must account for the ability of controllers to be aware of the process structure and inter-relationships among controllers. With more complex human-machine relationships in the future, a standardized systems view of both humans and software is needed. Safety of new systems will not be a matter of just reducing component failures or targeting human error rates, but instead managing a constrained dynamic process that does not allow or introduce hazardous behavior within the system. Indicators of mission risk should be based on each system's specific design and the implementation of human and machine roles within that architecture.

---

[6] These two airplanes are legacy MVs often used by the AF safety community as a reference.
[7] These terms, relating to adaptivity, are discussed in the next chapter.

### 1.3.2 Product Testing

Testing takes many forms; one is the act of simply putting a product or upgrade in the field and learning lessons from use. Someone who tinkers with their car engine to increase output will get their test results as they drive on public roads. Some products receive more formal testing in the form of lab or bench tests. Components or assemblies undergo rigorous and repeatable functional evaluations in a controlled setting. Sometimes entire systems, such as space satellites, are tested in a simulated environment (such as a vacuum chamber). The satellite must have undergone evaluation before experiencing the actual space environment, as its first venture to orbit is during field use. Inevitably designers will still learn some interesting lessons during its operations that could not be garnered from its terrestrial evaluation. Organizations that produce many different types of systems with varying degrees of complexity typically devote an entire enterprise to test (e.g., AFTC). Test professionals are responsible for building sanitized reproductions of certain real-world field conditions—as best as can be estimated—in order to evaluate aspects of the tested systems.

AFMC, headquartered at Wright-Patterson Air Force Base (AFB) in Dayton, OH, is the system-acquisition organization for the AF and delineates three stages of product life: a) research/design, b) test, and c) field-use/sustainment. There is certainly overlap among these activities, and sometimes items might get fielded quickly with minimal test, but the key principle is that most products go through the formal stages in some manner or another. Three enterprises (called centers) under AFMC manage these stages. AFRL, mentioned earlier, is the center that manages the first stage (research), while the Air Force Sustainment Center manages the third stage. *Development* typically refers to the first two stages.

AFTC, located at Edwards AFB, California, governs the second stage, formally termed *developmental test and evaluation* (abbreviated DT). During the DT stage of product life, the AF acquisition structure dictates a thorough verification by government testers of performance requirements and specifications of the system (or upgrade). This series of evaluations is performed with an emphasis on isolating new and unproven capabilities within a sanitized set of conditions that emulates (or blocks) the characteristics of the field deemed necessary to exercise the capability being evaluated. Typically after the DT stage, the system undergoes *operational test and evaluation*[8]—its first encounter with field-use representatives—to validate its utility in a battlefield representative ecology. Efforts to combine portions of DT and field evaluation are encouraged, so as to better steward taxpayer money as well as involve the field user earlier in the process [23].

AFTC manages the testing of many different technologies conducted in various geographical sites; the AFTC sub-organization responsible for the testing of air vehicles is the 412$^{th}$ Test Wing (412TW) collocated at Edwards AFB with AFTC headquarters. The

[8] In this thesis, "field use" is the formal term for everything that comes after the DT stage. The word "operations" refers to any active flight process whether it be during DT or field use. This is different than how the AF typically uses the term "operations," where it more closely resembles what this thesis calls "field use."

412TW also offers test-conduct and range support to non-AF organizations. Because of the proliferation of autonomy-capable military systems and concepts, AFTC is anticipating an increase in UVs flying at Edwards. The AFTC safety culture is interested in modernizing the safety-planning processes that affect airspace provision, range clearance, and flight-test procedures for these types of systems.

Developing a technology requires an elevated level of safety emphasis in the initial test stage before it is certified for field evaluation and use. This emphasis is called *test safety* [24]. Two driving factors for test safety exist: a) there are product configurations, instrumentation, techniques, and maneuvers that are unique to testing; and b) the inherent safety of the design itself is not yet proven—only modeled and simulated at best. DT not only evaluates performance but also serves as a risk reduction toward the field-use system safety.[9] Within AFTC, the general programs of *flight, ground,* and *weapons* safety—which complement tactics, techniques, and procedures (TTPs) in training and field implementation—are not enough. Test safety is thus additionally managed.

One example of risk reduction would be a new UV system going airborne for the first time. The sorties[10] performed to accomplish the initial flight envelope expansion follow rigorous test-safety procedures until the vehicle structure, propulsion, and flying qualities are proven. This maturity determination—verified consistency and integrity of the performance of the basic vehicle outputs—usually allows for the basic aircraft and flight controls to be managed sufficiently by the flight safety paradigm (or looser test-safety restrictions). Following that, the rest of the system's capabilities may be evaluated, governed by test-safety principles depending on techniques used. Also, it is possible that the two factors that lead to a test-safety level of scrutiny (unique evaluation techniques and the steady buildup required for risk reduction of the unproven design) can apply to the same test.

Testing autonomy-capable flight systems stretches the limitations of maintaining separate flight and test-safety protocols. Autonomy, as discussed in the next chapter, can be viewed from the perspective of decision loops and their components: sense, interpret, decide, effect [16]. A classic flight-safety paradigm is the “see and avoid” principle, required of all pilots (and their eyes) when visibility allows it regardless of whether they are following visual or instrument flight rules [25, p. 113]. This phrase is analogous to “sense/interpret/decide and effect.” Regarding inputs (sensing) and information processing (interpreting and deciding), historically there has been a pilot or aircrew in an airplane acutely monitoring the environment, maintaining an elaborate mental model of the situation, filtering potentially erroneous data, and in real time directing the actions of the vehicle. This traditional human role is considered even more critical when the onboard pilot is a test pilot. Flight testers are expected to rely on a wealth of experience to be able to manage complex testing activities.

---

[9] DT risk-reduction data also contribute to an airplane's airworthiness certification.

[10] A *sortie* in this thesis refers to a single flight by an aircraft or a single mission by a combined group of aircraft.

When testing UVs, the sensing, interpreting, and deciding functions are often all experimental. In the above risk-reduction example, the performance of the UV's flight computer, even before the basic vehicle performance is demonstrated, is already crucial to the flight safety of the system. Sometimes many immature technologies and/or novel software architectures must be tested concurrently. In UVs there is no human on the vehicle performing these functions in real time and place, and a human supervisor—at best—contributes to those functions from a distance. Even if thoroughly verified, the abilities of the UV (and remote human) to sense, interpret, and decide might still be limited compared to humans in MV cockpits. Fluctuating states in environmental awareness of the ground operators, sensor capabilities of the vehicles, and architectures of the flight team network form unique mismatches in relative maturities and consistencies of system functions. Robotic and autonomous flight technologies can develop so quickly that it can be difficult to characterize or test the performance of a single feature on a UV that is already considered safe to operate.

The current view during UV testing has been that system maturity determination comes from the *output*, or effect, of the vehicle onto the environment.[11] Military range safety guidance for UVs is based on legacy ballistic-weapons paradigms and conventional safety theory. Planners must calculate the probability that the physical vehicle will fail to operate, combined with the ground footprint the vehicle is capable of reaching if it crashes [27]. Airworthiness regulations extend these engineering reliability and maturity requirements to the UV ground control stations [28]. Unfortunately, safety issues stemming from complex behavior are not captured in this manner, and they transfer from the test to the field-use environment.

The progression to alleviate test-safety restrictions should be based on more than just proving maturity of the air vehicle's ability to effect consistent and predictable flight. When proving *system maturity* for the purpose of risk reduction, all four functions (sense, interpret, decide, and effect) should be verified. Deconflicting airspace and ranges to sanitize the testing of every function of a system becomes unwieldy when many of the functions are unproven. Additionally, sanitized testing conditions might not be ecologically valid when testing complex systems, since evaluators should examine entire decision loops instead of isolating components of the loop to verify their performance. "Safe to fly" and "safe to fly on a populated range with other aircraft nearby" sometimes are the same test. More targeted efforts must be made in DT to a) achieve ecological validity to the intended real world use, and b) qualify the maturity of the UV throughout all of its functions and not just the vehicle performance. An updated system-safety methodology based on STAMP can treat all system behavior as a control problem with decision loops. This would allow an approach that aligns well with autonomy concepts including use philosophies from the field ecology in early hazard analysis.

[11] The current operating instruction at Edwards AFB [26] differentiates between a UV system's maturity determination and type determination. Maturity is determined from the number of flight hours the design has and any available vehicle component reliability data (control surface actuators, airframe integrity, etc.) that can be used to infer the vehicle's capability to effect safe flight. Type is determined by the system's abilities to sense and react.

The acknowledgment of the blurred line leading from initial testing to fielding of autonomy-capable systems invites a proposal for a holistic integration of the various behaviors that can contribute to safety hazards into a complete system-safety approach. Hybrid DT/field-evaluation efforts already introduce a spectrum of safety risks as programs of various technology maturities operate in common test scenarios. Where DT testing is traditionally followed by a field evaluation to ensure ecological validity of the new system or upgrade, DT testing of autonomy-capable systems might lend itself to ecologically valid conclusions earlier in the development cycle because it directly tackles the safety issues that are and will be found in the field. An updated system-safety methodology could potentially be integrated broadly with MV policy as well.

### 1.3.3 Safety in Modern Systems

Hazardous system behavior—and the accidents that might result if the environment presents the opportunity—are the consequence of more than merely a chain of events or simultaneous critical failures. In modern sociotechnical systems and processes, hazards can be viewed as complex dynamic developments that include not only component failures, but also undesired interactions among parts of the system, unexpected (by the designer or policy maker) human/software behavior, and design and requirements flaws. STAMP is a top-down model that guides new system designers and post-accident investigators alike to consider a system as a hierarchical structure that requires appropriate controls between levels of function to ensure the system does not migrate to a hazardous state. The undesired migrations might occur over the length of a single operating process or along the lifespan of the sociotechnical organization as technologies and policies change [5].

STPA, a hazard analysis based on STAMP, works differently than a traditional hazard analysis in that the emphasis is always on the vertical structure of control and feedback. Instead of identifying multiple possible event chains that physically propagate to cause a top-level accident, the top-level accidents and hazards are defined first and then examined through the system's functional levels as safety constraints. Each control relationship in the hierarchy is analyzed for explicit causality of each hazard. In this manner, safety (viewed as freedom from accidents) can be traced to explicit hazard scenarios and mapped as a top level emergent property of the system. By referencing the design mitigations suggested by the hazard analysis, safety concerns can be considered with other top level requirements such as performance, cost and schedule. The concept of control loops also maps very well to autonomy decision loops.

AFTC presents an opportunity to apply STPA to test-safety management. Like many sociotechnical constructs that began in the last century and are coming of age with modern technology, flight test policies are a combination of many decoupled and derivative regulations and procedures that have evolved naturally and are applied in parallel to ensure safe operations (e.g., [29]–[34]). People are relied upon to do their parts during the planning and operating processes.

No single independent safety discipline (flight, test, range, etc.) can currently account for all process hazards. Likewise, accident and close-call investigations have a tendency to highlight traditional event-chain based traceability and root-cause

determinations without looking at dynamic interactions among components of the system and organization [35]. A systems-based technique for visualizing and discussing the interrelationships of multiple factors does not exist.

Recent efforts at the 412TW to combine general operations safety with test safety have resulted following the organizational changes that created AFTC. The 412TW safety office is currently using probabilistic event-chain hazard models and footprint calculators to design test-range safety constraints for UVs. The airspace managers are still using spatial and temporal deconfliction between UVs and MVs even as available airspace diminishes with the proliferation of products that need to be tested. Current planning practices do not adequately to model the multi-dimensional and inter-organizational control structure at Edwards; as such, the test community is looking for a modern methodology to better assess complete system safety, particularly with respect to autonomy.

MIL-STD-882 lays out the Department of Defense (DOD) product safety road map: identify hazards, perform Preliminary Hazard Analyses (PHA), map subsystems to top-level requirements, and eliminate or reduce risk throughout the stages of system development and acquisition. 882 reiterates the objective of DT safety planning, which is to "eliminate [or reduce] the hazards for both the system and the test [unique] events" [11, p. 82]. Thus, testers must both evaluate acceptable system safety (as defined by the program managers) and account for risk to test personnel, range/support personnel, and the public.

Typically the local safety process employed during DT consists of identifying hazards, a form of PHA called test-hazard analysis (THA), and risk reduction efforts informed by the experience and best practices of senior test personnel. The remainder of the MIL-STD guidance—mapping system and subsystem-safety requirements, functional hazard analyses (FHA), and inter-operability safety assessments in the larger field-use ecology—is left to the program managers to implement throughout the remainder of system development. System safety encompasses more than just the reduction of risk on one vehicle during DT. With some formalization applied to the method of defining accidents and hazards, followed by a modern hazard analysis approach, the potential exists for DT to enhance the system-safety assessment of the product.

General risk-management initiatives are also emphasized in the AF to reduce the likelihood of hazards during general operations. Training on crew resource management (CRM) is presented regularly to encourage the appropriate use of communications during a mission and the recognition of mentally deficient states in oneself and one's teammates [36]. Also, right before a sortie, all crew members must fill out a questionnaire regarding several factors that studies have shown to be correlated to higher accident rates (e.g., "how many hours of sleep did you get last night?") [30]. These initiatives, however, focus on the human operators and on maximizing their potential to break a mishap chain—almost as if right before an accident they become outside overseers, able to function without the limitations of the system [37]. This method of risk management is not specific to the detailed system design, the events of the specific operating process, or the organization that manages the program.

As mentioned in Section 1.3.1, the three traditional problem-solving areas in the UV domain have been human engineering, software algorithms, and process regulation

and/or policy. The AF could benefit from a common safety strategy that translates across all stages of product development and treats risk mitigation as a control structure. STAMP treats both human and software behavior as complex phenomena that cannot be considered independently in reducing system hazards. STPA goes beyond suggesting better policy by recommending designs and constraints that allow controllers, both human and artificial, at all levels to observe feedback (sense) and implement effects (act) with appropriate understanding of the process states and options for action (interpret and decide).

## 1.4 Research Methods

The purpose of this research is to improve the ability to assess system safety during developmental product testing (DT) and standardize the applicability of hazard findings between the design and field use of the product. The research first extends the methods in STPA to analyze the behavior of humans in complex systems. It then formalizes test as a dedicated stage in STAMP. The research then evaluates a test-safety analysis in AFTC for a UV use case. The implications span beyond the AF application, and the goal is to create a framework that testers can use to conduct safe and effective evaluations to make determinations about inherent system safety.

### 1.4.1 Tasks

The first research objective is to extend STPA to better examine human controllers in the hazard analysis. I performed this work at MIT. In order to understand the existing guidance for analyzing human controllers in STPA, I reviewed past STAMP models and analyses for format and content. In order to extend the STPA analysis, I expanded the detail for examining humans as well as any general intelligent controller. The extended analysis is called STPA-RC (refined controller-analysis). I added granularity to the controller mental-model analysis, motivated by previous work by Leveson on software and automation modes and Rasmussen's emphases on documenting high-level system goals and values [8], [38]; I detailed fundamental human-engineering considerations into the STPA analysis to account for guidance found in existing Human-Systems Integration (HSI) practices [39]–[42]; I also developed a method to summarize the influences to controllers that evolve prior to a system operation, to include engineering configurations and controls as well as socio-organizational stimuli [43]. This research produced a tool called an Explicit-Influence Map (EIM) that complements the new guidance by tracing the explicit policy and planning products that inform controllers in a specific industry.

In order to examine the utility of STPA-RC, I compared results produced with it to results from two previous STPA analyses performed on an FAA Next-Generation air traffic control (ATC) concept called In-Trail Procedure (ITP). Results were tabulated for all three evolutions of STPA and contrasted along the logical partitions of the updated method. STPA-RC encourages engineers to iteratively update a system's model and safety planning documentation based on questioning assumptions about feedback and communications between controllers.

The second research objective is to provide a common framework for test-safety planning that addresses both the safety of the test process and inherent system safety. I accomplished this work with several visits to Edwards AFB over a two-year period, supported by Draper Laboratory. I am an AF acquisitions officer with access to key personnel within the AF research and test enterprises. This allowed for the collection of SME perspectives during research to formulate notional organizational products that would normally require many members of an organization to produce cooperatively. In order to learn the sociotechnical structure of a professional test enterprise, I performed a thorough organizational and policy review for the 412TW, supported by fact-finding interviews, including in-depth contextual inquiries of test practitioners [44]. I created an EIM for the AF 412TW, using a freeware concept-mapping tool called Visual Understanding Environment (VUE).[12] I incorporated all applicable policies, academics, planning tools, and resources into the map to demonstrate the usefulness of the product compared to other methods of tracking those influences and policies.

In order to model the systems view of testing, I produced a new generalized example the STAMP organizational control structure to include a testing stage. I developed new inclusion criteria for practitioners to use when developing their industry-specific safety-control structures. The new structure accommodates both safety of the test process and inherent system safety. The safety of the test process is argued to be a special case of inherent system safety, and the tasks and communications required to analyze test safety and produce risk assessments are highlighted via the test-management control loops shown in the test stage in the generalized example. System-safety tasks and communications are highlighted via multiple organizational feedback across each stage and include: certification information for users and regulatory agencies; hazard analyses, control models, and design rationales and assumptions; and problem reports.

In order to further examine safety of the test process, I developed a format for performing and documenting STPA for testers. I used STPA to analyze a flight test of a real experimental product involving an autonomous wingman flying next to a human-piloted lead aircraft. This project was flown at Edwards to demonstrate the behavior of autonomous flight algorithms. I compared mitigations produced by the STPA format to the results of the local traditional THA-based report. I also compared subjective aspects of the two reports by administering a research survey to flight test professionals at the 412TW. The subjects were given the safety documentation resulting from of both the traditional and STPA analyses, and they were queried for preferential opinions on the intelligibility, informativeness, and implementability of the two methods.

### 1.4.2 Thesis Structure

Chapter 2 gives a brief overview of systems theory and human engineering. System safety is introduced, followed by a summary of STAMP and STPA as modern engineering approaches to system safety. The Air Force, its autonomy philosophy and research thrusts,

[12] Tufts University, © 2013. http://vue.tufts.edu/

and its acquisitions and product-development organization are introduced as an application for this research.

Chapter 3 describes the methods and results of the first research objective. A discussion of the visual format for illustrating the safety control structure STPA tables is shown with a discussion about information ties between them. STPA-RC is presented by modifying the existing controller analysis, introducing new visual aids, discussing the implementation of additional guidance where appropriate, and introducing some terminology as required to discuss the new concepts and support the remainder of the thesis. The EIM is introduced as a new planning tool, and policy is discussed. The ITP example then shows a practical comparison between STPA analyses with and without the extension.

Chapter 4 describes the methods and results of the second research objective. It discusses DT philosophy, reviews product and operations safety practices in a complex organization like the AF, and introduces AFTC as an established instance of a product testing enterprise. The EIM product for the 412TW is discussed. A new proposed generalized example of a STAMP organizational control structure is presented, with a discussion on the additional fidelity provided by a dedicated test stage and the information communicated within and between stages. A real 412TW test project is analyzed for safety of test using STPA, with results compared to the traditional approach.

Chapter 5 summarizes the research and findings, discusses implications and limitations, and makes recommendations for current applications and future research.

# Chapter 2

# Background

> "It is possible to fly without motors, but not without *knowledge* and skill."[1]
>
> —*Wilbur Wright*

This literature review is divided into two main parts: system safety and Air Force systems. System safety broadly discusses systems theory and the different historical viewpoints of the role of humans in systems, ending with an introduction to a modern system-safety model that serves as the basis for work done in this thesis. This material gives the reader background for understanding the research I discuss in Chapter 3, which is also needed to fully appreciate the research in Chapter 4.

The Air Force material further extends the motivations introduced in the previous chapter, introduces the concept of developmental test and its place within system development, and discusses autonomy as a looming interest area for aerospace research. This complements the human discussion as a basis for making improvements in the ability to incorporate humans into hazard analyses. Furthermore, the more in-depth discussions about testing in Chapter 4 are easier to follow given an understanding of the material in this chapter.

## 2.1 System Safety

Checkland states that the distinction between science and engineering occurs with the creation of artificial objects by humans, and that engineering is interested in "action directed to a defined end" [45, p. 127]. As man-made systems[2] began to grow more elaborate beginning in the middle of the last century, the intellectual complexity required to design, build, operate, and manage them increased immensely. The growing potential of software as well as new discoveries in human sciences regarding cognition resulted in a wealth of laboratory research throughout various disciplines to address Human-Systems Integration (HSI). However, neither software behavior nor human behavior can be

[1] In a letter to Octave Chanute, 13 May, 1900 (emphasis mine).

[2] The National Aeronautics and Space Administration (NASA) Systems Engineering Handbook defines a system as: "The combination of elements that function together to produce the capability to meet a need. The elements include all hardware, software, equipment, facilities, personnel, processes, and procedures needed for this purpose" [46, p. 275].

completely explained by physical laws when they are coupled in complex systems and within dynamic work ecologies. System designers often still treat human operators as overseers of systems instead of an integral part, or they compare humans and machines at the level of basic performance capacity and thus design machines solely to compensate for apparent human inefficiencies [47].

The scientific method began to be complemented by *systems theory* in the 1940s precisely because unrestricted sciences[3] exhibited phenomena that were not completely explainable in the language of their basic physical principles. There were properties of systems that were of interest to humans (for observation and/or manipulation) but that were not definable—or really existing—if the system were only viewed as a collection of simple components. *Safety* is an example of such a property. It cannot be attributed to atoms, molecules, materials, or even machine components. Safety is a top-level property of any natural or man-made system in which people have identified goals and acknowledged the types of losses that should be avoided.

Engineering has embraced the findings of scientific research when necessary, historically favoring practical applications before theory and methodology [45]. When technology became more complex during the last century, design and management methodologies appeared with the intent of organizing complexity via the creation and optimization of models and task hierarchies, and systems engineering was established. Systems theory had some influence in this stage, as it was itself in its early growth phase in the sciences. Safety was one area that benefited, motivated by the systems challenges of designing intercontinental ballistic missiles in the 1950s [48]. Safety engineering was at the time an integral part of system management, and the process would yield *systemic* approaches to hazard mitigation. Progressing into the later century, however, the scientific method retained a greater influence over the less-established systems theory. Discipline-specific engineers continued to rely on reductionist-science approaches to guide design, and safety planning evolved into a function of tracing the simple physical relationships of low-level components. Safety engineering as a discipline retained the *systematic* qualities (model building and optimization) of systems engineering and management, but as a modern discipline it has lacked systemic qualities [5].

Stakeholders working within or supporting sociotechnical systems need to treat safety as "an emergent property of the organized complexity" [49, p. 6]. Managing safety by focusing exclusively on component failures (e.g., reliability, human error) has limited effectiveness and has even created unintended consequences. Because top-level system properties and goals (Checkland's "defined ends") cannot be attributed to the system's reduced bare components, safety must similarly be understood starting at the top of the system. System-Theoretic Accident Model and Processes (STAMP) is an accident model developed at the Massachusetts Institute of Technology (MIT) by Nancy Leveson [5]. It can be used, through various methods of application, to trace systemic behavior throughout a system's entire lifecycle. This makes it a powerful model that can complement the

[3] Biology and sociology, for example (as opposed to physics and chemistry, in which a restricted range of phenomena are of interest).

systematic practices that exist in modern systems engineering and management to ensure safety.

### 2.1.1 Systems Theory

One of the large contributions to the scientific method was by Descartes, who argued for discovering simple natures to explain seemingly complex phenomena. This idea of *reductionism* is a fundamental underpinning of classic science, along with *repeatability* of findings and *refutability* of hypotheses. Descartes argued for a rational view of the world that could be explained by the sum of its observable basic constituent mechanisms rather than by an external purpose (teleology).

As the levels of complexity increase, so do the limitations of scientific reductionism.[4] In social sciences, there are limited opportunities to gather enough experimental data to produce unequivocal results, as well as difficulties generalizing beyond the internal validity of an experiment. Additionally, investigations into human behavior bring the unavoidable possibility of the observer attributing unique meanings to the data collected, and predictions about future behavior can affect the behavior itself. Finally, when research is intended to improve aspects of sociotechnical systems, the challenges themselves are difficult to define, and unique meaning and purpose *must* be applied by human stakeholders in order to identify the problems [45].

In biology, von Bertalanffy borrowed the concept of *holism*, a term coined by Smuts, to offer a complementary viewpoint to reductionism [50]. He viewed natural systems as more than a collection of parts and solidified a concept of organized complexity in which there exists a hierarchy of levels of organization. Every level is more complex than the one below and exhibits *emergent* properties that do not exist and cannot be explained in the languages of the lower levels. Thus, a different level of description is required for every level of complexity. He insisted that emerging ideas within the various scientific fields[5] could be captured by a general systems theory [51].

The systems approach, seen as a meta-discipline, is meant to complement the scientific method, which aims to acquire testable and refutable knowledge of the universe.

---

[4] Checkland describes scientifically acquired knowledge as "the best description of reality that we have" in order to provide predictive power of natural phenomena "at that moment in time" [45, p. 50]. Reductionism began to show its limitations within the field of biology in the early 1900s [45]. The rich variety of its many observable phenomena had exhibited complexities beyond which the scientific method could provide adequate descriptions. Chemistry could provide an account of the mechanisms for some biological phenomena, but it couldn't explain the existence of or the answers to biological questions. In other words, fully controlled experiments crafted in the language of chemistry would not be able to observe and predict every chemical interconnection of a biological process or answer why a process at the level of a biological organism behaved as such. It was eventually realized that, similarly, physics might explain mechanisms of some chemical behavior and biology could explain some aspects of human behavior, but each level could not completely explain away the one above it without the method becoming unmanageable.

[5] The major scientific fields of the time (i.e., physics, chemistry, biology, psychology, sociology, etc.) could even be viewed as displaying progressing emergence with respect to each other.

Systems theory recognizes that systems might be natural, designed, or human-activity.[6] While the principles of scientific observation are always valid for collecting data, the appropriate questions of *what* needs to be discovered and *how* to explain it in language appropriate to the level of complexity of interest fall to systems theory. Designed systems and human-activity systems in particular exist for a *why*, a purpose,[7] and thus a reductionist approach alone limits the types of useful explanations for real human problems.

Besides emergence and hierarchy, systems theory embraces the idea of *communication and control*. This began with an acknowledgement by von Bertalanffy that *open systems* have a *boundary* with the environment, and that in order to maintain order, a set of processes must exist to regulate the exchange of materials, energy, and information between the system and the environment. This concept, which began to be realized from the observation of natural systems such as organic cells, was extended to designed systems and human-activity systems by Wiener in the 1940s through the study of *cybernetics* [52].

Wiener is thought to have made the greatest individual contribution to systems theory [45]. He coined cybernetics as a discipline which would study control and communication theory for all systems, natural or not. Based on both existing biological and control-theory principles, cybernetics treated *feedback*—the transmission of information about a process to the process-controller—as the underlying notion for organized activity. His work was motivated by problems with military fire control radars during World War II, in which anti-aircraft weapons needed to track enemy airplanes and predict their future positions in correlation with the ballistic opportunities of the defensive guns. Although the term is not often used today, cybernetics as a qualitative philosophy of technology contributed to many disciplines still in use such as systems theory, information theory, and computer science [53].

Control is needed in complex, open systems because there is uncertainty in both the environment and the observable characteristics of the system. While a closed system will settle into a state of equilibrium, an open system will migrate into another state if not for regulation of its processes. Hard engineering paradigms, while able to design mechanisms and controls (syntax), might be too brittle to preserve the meaning (semantics) and intent of more complex human-activity systems. It is for these problems that Checkland extended systems theory by considering soft-systems approaches that help stakeholders determine the purpose of their design, engineering, and problem-solving endeavors [45].[8]

---

[6] Natural systems have origin in the forces and process which characterize the universe; designed systems may be physical (hammer) or abstract (poem) and originate via conscious design to fulfill a purpose; human-activity (sociotechnical) systems consist of multiple self-conscious activities coherently related and continuously evolving as a whole towards a purpose; [45].

[7] Teleology (which Descartes discounted when arguing for reductionism) implies a system purpose extrinsically designed by an agent; teleonomy (quality of apparent purposefulness) is a more neutral term and can apply to natural systems as well.

[8] Sometimes there is no well-defined goal, and instead there exist problematic situations which require inquiry by stakeholders. The model of the problem situation becomes the system for analysis, and the design, development, testing, and implementation of solutions are themselves human-activity systems.

#### 2.1.1.1 Use of Abstraction and Models

Model building is a way in which humans use abstraction[9] to interpret their surroundings. Models can be used to represent curiosities and perceptions of the world, understand human participation within natural systems, or guide the creation of designed systems. Human-activity systems typically involve all of these activities (identifying, using, and designing/implementing). Box said, "all models are wrong, but some are useful" [54, p. 424]. Models of the physical world are constrained by their assumptions. Typically, high fidelity models of complex phenomena can produce elaborate outputs given expected inputs, but the ability to predict accurate results is limited by the "astounding variety and richness" of those phenomena, which produce uncertainty in the assumptions [45, p. 162]. Complex systems are better understood and influenced when the model is abstracted to focus on the *problem* to be solved.[10] Details that are not important to the problem for the level of complexity in which it is defined need not be in the model for it to be useful.

When modeling human-activity systems, many people must be involved to define goals, assess problems, and provide expertise into the details of the design as they relate to the functions of the system. Star and Griesemer encouraged the use of a shared model[11] to communicate the representations of people's ideas and perspectives, using a format "plastic enough to adapt to local needs and constraints of the several parties employing them, yet robust enough to maintain a common identity across [disciplines]" [56, p. 393]. Models like this allow each person's local understanding to be put into the perspective of a shared activity, linking disciplines together to collaborate on a problem [12], [57].

Systems thinking includes the use of models to guide communication among planners and to involve stakeholders. Particularly when displayed visually, models can be more useful than mere conversations or planning reports. As the group builds and refines the model, they take ownership of it and get increasingly specific about the problem they are trying to solve, including the questioning of assumptions. Thoughts and perspectives can be modified and shared, leading to new insights. Individual mental models may be inaccurate in the context of the organizational objectives, but group interpretation makes models useful to the objectives of the system. Meaning is created through the interaction among stakeholders.

### 2.1.2 Humans in Systems

It is possible to adopt a system-theoretic viewpoint of human behavior so that the human role in systems can be understood, managed, and improved to achieve the goals of an organization. Safety is one of those goals. Part of this thesis research was performed to extend system-safety techniques to refine the analysis of humans in complex systems.

---

[9] An abstraction is a structured product of the conscious mind (thoughts or concepts) that exists to serve a human purpose [45]. It can be used to generalize theories based on perceived facts, or represent ideas in a more general form than their physical manifestations. Mathematics, music, art, and verbal language all use abstraction in this way.

[10] In a system-dynamics approach (not used in this thesis), the specific characteristic stakeholders want to control would be called the *reference mode* [55].

[11] Their term was *boundary object*, which is not used here.

Thus, an overview of the history and development of human engineering is presented here to put the modern viewpoint into context before the extension of the system-safety approach is presented.

Many volumes exist to summarize the history of human factors (HF), human-systems engineering, HSI, and other disciplines concerned with the relationship of humans and technology [58]–[60]. Those disciplines in their collective efforts serve to apply biological and behavioral sciences to understand interactions among human and non-human entities and design systems to optimize human well-being and general mission performance. Until recently, approaches to improve human-machine systems have divided the problem into comprehending machine-performance capabilities separately from human-performance capabilities, logically isolating the two via interfaces. Humans are often considered overseers or supervisors existing separately above machine systems, and this reductionist viewpoint permeates many industries today.

Since the late 1800s, various engineering fields began devoting some or all of their efforts to integrating humans and artificial components for the purpose of work. Taylor applied engineering methods to the labor force and developed principles that would eventually contribute to the development of industrial engineering [61]. He advocated for fitting laborers to tasks, in the name of efficiency, and this was done by rigorous user-selection and pay-incentivization mechanisms. The first modern discipline dedicated solely to human engineering, *ergonomics*, initially handled the concept of the workspace as such. As academic interest in human-task relationships increased, the mindset would evolve instead to that of fitting the tasks to the operators.[12] Ergonomics first experienced that shift, and subsequent human-engineering endeavors have gone through similar transitions [62].

Physical ergonomics is concerned with human anatomy and making physical work efficient. Anthropometry, the measurement of individual physical traits (e.g. a person's sitting height), can be used to determine population percentiles to fulfill work that requires particular dimensions or characteristics for operator safety and/or efficiency. The military today still levies anthropometric constraints onto many of its specialties, and they have contributed much to the field [63]. Ergonomics has also taken advantage of research in human motion (kinematics and kinetics), pioneered by the Gilbreths in the early 1900s [64]. This knowledge led to design optimizations to make work tasks achievable by a larger percentage of the potential operator population. Eventually design engineers began designing better workspaces in terms of safety and efficiency.[13]

The ability to display various technical data to operators through workplace interfaces became a fascination for designers, and coupled relationships between humans and controlled processes were acknowledged in the light of the cybernetics movement. Flach, with others, discusses three different viewpoints to human-machine systems:

---

[12] Many advanced industries including aerospace still select operators based on rigorous criteria, still influenced by vestiges of Taylorism.

[13] Areas of improvement included optimizing mechanical geometries and avoiding overstraining muscles and joints during physical work, better arrangement of components in the workspace, and accounting for environmental factors (such as climate, lighting, and noise) as well as perceptual and physiological limitations [65]–[67].

technology-centered, user-centered, and ecological [62], [68], [69]. Each one has been a strategy for understanding system behavior and guiding design, and they represent a progression of thought that parallels the growth of systems theory in the sciences in general. The first two viewpoints are reflected in the following two sections, and the third viewpoint is discussed in Section 2.1.3.1.

#### *2.1.2.1 Technology Centered Viewpoint of the Human*

Into the middle of last century, research began to attempt to understand the differences in the abilities of humans and machines. Fitts—considered by most today to be the father of HF—following investigations of human performance during aircraft control, formulated his famous "MABA-MABA"[14] list, reproduced as follows [70]:

- Humans surpass machines in these abilities (Fitts, 1951)
  - Detect a small amount of visual or acoustic energy
  - Perceive patterns of light or sound
  - Improvise and use flexible procedures
  - Store very large amounts of information for long periods and recall relevant facts at the appropriate time
  - Reason inductively
  - Exercise judgement
- Machines surpass humans in these abilities (Fitts, 1951)
  - Respond quickly to control signals and apply great force smoothly and precisely
  - Perform repetitive, routine tasks
  - Store information briefly and then erase it completely
  - Reason deductively, including computational ability
  - Handle highly complex operations (i.e., do many things at once)

Fitts, as cited by Winter, later admitted he "fell into the trap of trying to make a list" that was "trivial and somewhat misleading," but the idea of functionally allocating tasks between humans and machines persists to this day [71, p. 8]. The view of humans—albeit having natural performance limitations—filling the role of active overseers of machine systems began to form the rough-order cybernetic approach to engineering design.

In systems that were tightly coupled[15] between human operators and machines, manual control was investigated quantitatively using analytical control theory. Human limitations such as neuromuscular lag and time delay were characterized into approximating transfer functions in order to optimize the handling qualities of various

[14] "Men are better at / Machines are better at."
[15] This refers to instances in which the human controller is actively involved and making corrections via the use of continuous inputs to achieve a setpoint or reference state.

machines and vehicles under human control [72]–[74]. It was also realized that humans have a capacity to alter their manual control strategy to produce a combined human-machine transfer function capable of consistent dynamics within reasonable limitations [75].

Outside of manual control theory, and despite the understanding from systems theory that dynamics can be controlled with closed-loops between hierarchical levels, the primary focus of control engineering as a discipline was on machine components and not on human-machine interactions. Advances in software eventually automated functions that were formerly in the realm of manual control. Most industries (some sooner than others) have begun to treat humans as primarily serving a supervisory role instead of that of a continuous controller.

Sheridan, through his work in telerobotics, acknowledged the increasing range of supervisory relationship types between operators and the systems they oversaw [76]–[78]. He introduced the concept of *levels of automation*, which details the tasks that a computer or machine performs with respect to a human operator. The following list is a reproduction of the original, which has been extended and expanded by others in subsequent research [79], [80]:

1. (Lowest Level of Automation) – Computer offers no assistance; human must take all decisions and actions
2. Computer offers a complete set of decision/action alternatives, or
3. Narrows the selection down to a few, or
4. Suggests one alternative, or
5. Executes that suggestion if the human approves, or
6. Allows the human a restricted veto time before automatic execution, or
7. Executes automatically, then must inform the human, or
8. Informs the human only if asked, or
9. Informs the human only if it, the computer, decides
10. (Highest Level of Automation) – Computer decides everything, acts autonomously, and ignores the human

As more tasks transitioned from a manual to a supervisory relationship, design scrutiny would inevitably be given more to the machine component, focusing on optimizing its performance and displaying it to the human. Autopilots, for example, were able to perform much more optimally and with a shorter time constants than human pilots for various tasks such as maintaining altitude or other flight parameters [62]. However, this technology-centered viewpoint led to some challenges.

First, artificial controllers are responsive, accurate, and stable (and typically greatly outperform humans attempting the same task), but this is only the case when global assumptions about their environmental context are satisfied. If the assumptions change and the controller does not have a corresponding range of control laws or settings with which

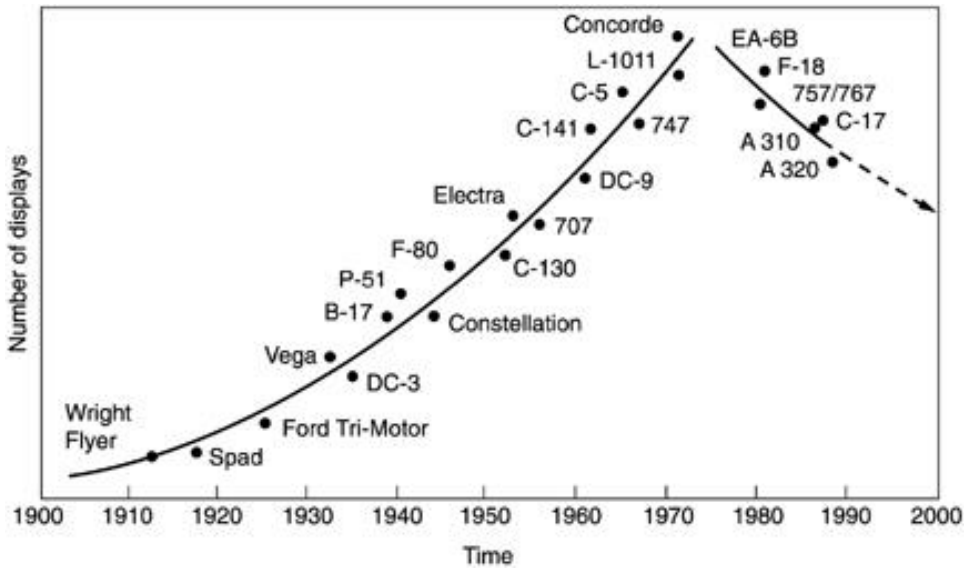



Figure 2-1. Historical Trend of Cockpit-Display Densities [85]

to compensate, it is a *brittle* controller and will become unstable. An example would be an autopilot that cannot integrate parameters appropriately for control at high altitudes [81], [82]. In cases of environmental uncertainty, a level of robustness and creativity is demanded of controllers. Humans, who are able to dynamically adapt their control strategies as discovered in the manual-control research, would be required to recognize the deficiency of the technology, take over the task, and impart stability using creative means. The capability for the human supervisor to sense brittle control and take over is not always straightforward to implement in a design, and it is made more difficult when the human-machine relationship is executed from a distance, as is often the case with tele-operated or remote-presence concepts like unmanned vehicles and robot-assisted surgeons [83].

Second, maximizing all technological capabilities (something that enthusiastic engineers strive to do) to create a suite of advanced tools for the human overseers can backfire if human operators become overloaded with information. Control displays might indicate readings for every possible monitored parameter, a characteristic termed *single sensor, single indicator* (SSSI) [84]. Figure 2-1 shows an example of display proliferation in the last century [85]. Once information-processing capacities of humans came into focus (discussed in the next section), the average number of displays began to decrease again.

Third, and related to the previous challenge, is the issue Wiener called *clumsy automation* [85]. During periods of low task density, automated functions would be such that human supervisors would stop paying attention to the system. Yet, when task density increased, the proliferation of displays and poorly balanced division of workload between humans and non-human components would actually increase the mental effort required by humans to manage both the tasks and the allocation of responsibilities. The nuclear accident at the Three Mile Island plant in 1979 served as one of the great prompts for systems designers to consider that operators might not be understanding what their displays were telling them about the system state [86].

Sarter and Woods, with others, introduced the concepts of *automation surprise* and *mode confusion* [87], [88]. Surprise refers in general to the component(s) under the supervision of a human operator not behaving the way the operator expects. A classic example would be when an airplane autopilot disables itself because its environmental assumptions or control limits are exceeded, but the human pilots continue to believe it is switched on until the airplane almost flies into the terrain. Mode confusion occurs when the operator cannot form a belief of the system state at all due to ambiguity in the way they are receiving data from their sources of feedback. Woods and others insisted that engineers need not automate everything possible nor follow the MABA-MABA mantra of functional allocation literally. Doing so would overcomplicate the system, and the focus should be on the coordination between humans and machines [89], [90].

Such a systems viewpoint of hierarchical controllers, feedback, and communication aligns with cybernetics concepts, and it has not gained momentum until recently. Leveson recommended six areas of focus for designers to prevent mode confusion without oversimplifying the system: interface interpretation, consistency of system behavior, knowledge of indirect mechanisms behind mode changes, authority limits, appropriate feedback, and minimization of unintended side effects [38], [91]. However, there is a more prominent movement based on engineering psychology that avoids the challenges of technology over-focus by instead treating the problem as an input-only, open-loop approach targeted at the human mental capacity. This user-centered perspective is discussed next.

#### 2.1.2.2 User Centered Viewpoint of the Human

Engineering for the activities of the human mind became absorbed into general human-engineering efforts, which were formalized into HF in the mid-century. In experimental psychology, *behaviorism* was the dominant paradigm in the early 1900s [92], [93]. Introspection or focus on the mind's awareness were not considered, only the mapping of responses to stimuli. Early influences of this mindset on HF reflected this. Fitts's Law, for example, showed that the movement-track time of a response was a function of the ratio of the distance to the target divided by the size of the target [94]. This supplemented Fitts's previous work on human errors and limitations. The technology-centered approach, emphasizing the selection of operators based on high mental and physical aptitudes for complex work domains, began to give way to the consideration of the variance in human capabilities. HF experienced a shift in focus, as ergonomics had previously, from selecting the right man for the machine to instead designing the machine for the man.

Human models based on behaviorism predicted reaction times to simple stimuli. Factors like the modality of the stimuli (i.e., visual versus aural), intensity of stimuli, and expectancy of stimulus signals were shown to have predictable effects on reaction times. By the 1960s, Simon showed that when multiple stimuli and responses are possible, if the location of a response action (e.g., the button that needs to be pushed) is proximally located to the corresponding stimulus the response time is faster [95]. It was around this time that psychology began a shift to *cognitivism*, which opened up the human mind to study. This allowed mental functions such as information processing and problem solving to be incorporated into HF analyses. This was a step forward for human-engineering research,

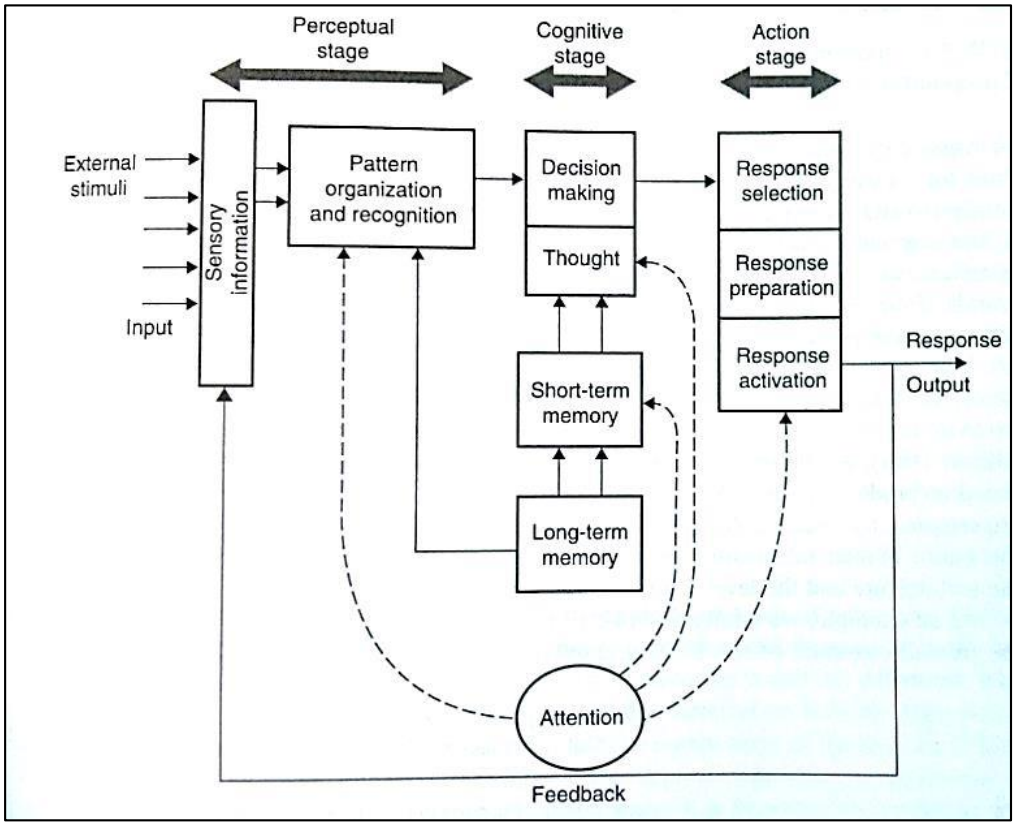



Figure 2-2. Human Information-Processing Model [59]

even if the reductionist mindset established by earlier periods of human study carried through.

Engineering models, borrowed from information-processing and signal-detection theory, were modified for human behavior starting in the 1950s. This mirrored progress in communication systems, control systems, and computers after World War II [96]. Reaction-time experiments grew more elaborate, and mathematical approximations were fashioned to estimate human rates of information flow, stimulus discriminability capacity, and rule-based response performance. Chronometric methods were introduced (e.g., additive and subtractive tasks, psychophysiological measures) in order to form models of discrete information-processing stages in the human mind. A version by Wickens is shown in Figure 2-2 that includes explicit stages for perception, cognition/world building, and response determinations. Also represented are types of human memory[16] proposed by Atkinson and Shiffrin in the 1960s and later expanded by Baddeley and Hitch in the 1970s [59], [98], [99].

[16] These include sensory stores, short-term or working memory (along with the visuopspatial sketchpad and phonological stores), and long-term memory. One of the more famous cognitive rules of thumb is Miller's Law, which states that the average person's working memory can hold 7 ± 2 objects [97].

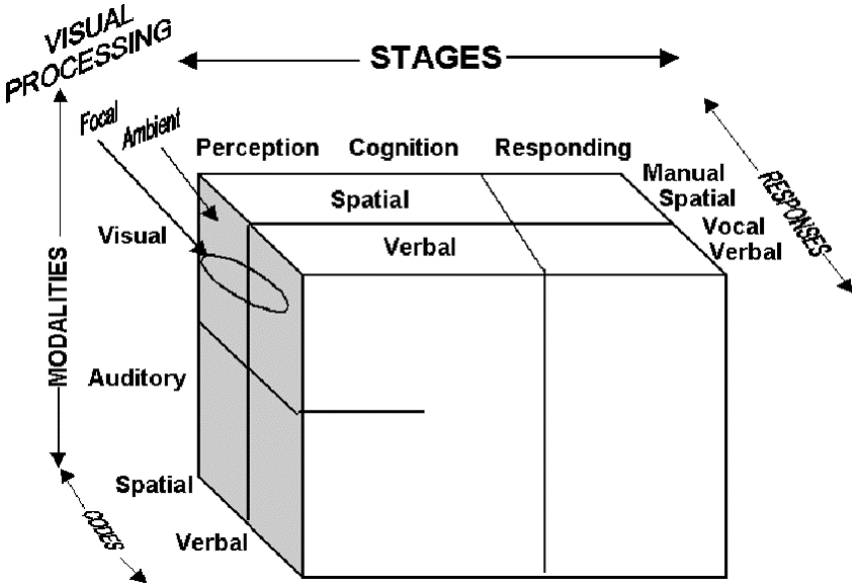



Figure 2-3. Multiple Resource Theory [59]

Under various conditions it was found that humans performing multiple tasks (in time and/or space) would encounter limitations in their ability to process and respond. Accuracy and/or timeliness were degraded compared to the performance on singular tasks. Different explanations were proposed for bottlenecked performance, including various perceptual filtering theories and the psychological refractory period[17] [100]. The idea that mental resources were being shared and demanded by multiple tasks as well as different functions of mental processing was represented by various resource functions and operating characteristics. Wickens also developed the Multiple-Resource Theory to simultaneously account for the various dimensions that previous experimentation had shown individually to affect the sharing of mental resources. Shown in Figure 2-3, this model is meant to be interpreted in the fashion that if a line separates two volumes, they do not share mental resources. This model has influenced many studies on distractions and multi-tasking [59], [101].

*Attention* was introduced as the finite mental fuel that could be allocated to multiple tasks, including the executive control function required to switch focus between tasks. Interest was renewed in the Yerkes-Dodson Law, originally developed in the early part of the century from a behaviorism standpoint [102], [103]. That research had shown that an empirical relationship exists between arousal and task performance. Work performance increases with increased mental load, but only to a certain point at which performance decreases. A representation is shown in Figure 2-4. Cognitivism theories contributed the concept that attention itself is malleable, and that while at higher arousal human performance and memory become resource-limited, at lower levels of arousal there is also

[17] Studies determined that the processing meta-stage of response selection and activation uses a limited mental resource that is separate than the resources for the previous meta-stage (perception and cognition). Tasks could be introduced in a certain order that resulted in the response selection of a later task delaying the response activation of an earlier task.

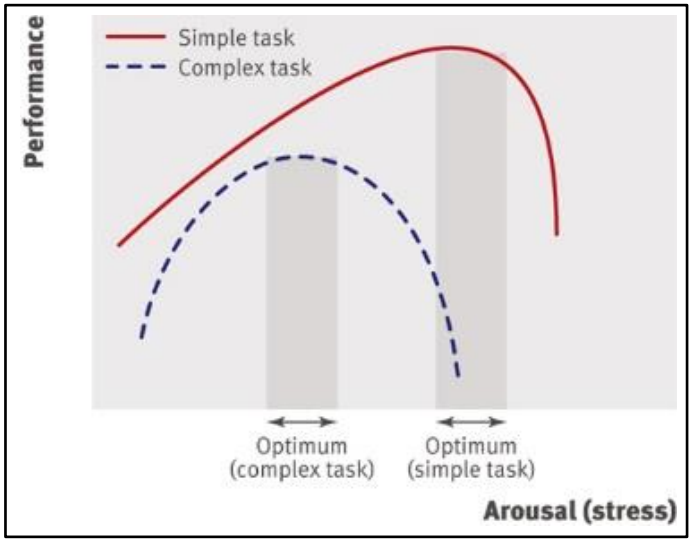



Figure 2-4. Yerkes-Dodson Law (http://www.blog.theteamw.com)

a performance degradation due to a naturally reduced attention capacity. This research began to tackle the challenges of clumsy automation discussed in the previous section, and there are many modern efforts targeted at understanding the human capacity for vigilance during multiple prolonged tasks at various levels of arousal [104].

These concepts were developed into engineering practices. Attention was distributed into a taxonomy that divided *sustained* attention (where vigilance decrements occur) and *selective* attention. In selective attention, multiple information displays are available to a human operator, and the switching between displays could be *endogenous* (by choice of the operator) or *exogenous* (by form of a cue or distraction). The chosen display would then be the subject of *focused* attention, where studies in perceptual phenomena, such as change blindness,[18] aim to optimize the placement and density of information. These techniques grew into static and dynamic display principles based on work by Gestalt and Fitts as well as stimulus-response and proximity-compatibility methods adapted from prior ergonomics principles [59].[19]

Cognitive-engineering constructs were also developed [105]. Mental w*orkload* was considered the mental-effort cost of accomplishing tasks. If attention was the available fuel, workload was the fuel requirement. Various experimental measures were designed to assess human workload via objective and subjective techniques.[20] *Situation awareness*

[18] This is a phenomenon in which so much information is put on the same display that an operator, focusing on some of the information, misses an important (and otherwise obvious) indication within the same visual focus.

[19] Examples of these principles can be found in the placement of control knobs for stoves, problem indicators for aircraft engines, and more complex instruments that must portray representative domain information in a readable and understandable format, without oversaturating the perceptual abilities of the human.

[20] Objective techniques included primary and secondary task paradigms to measure the effects of loading, as well as psychophysiological measures such as blink rate, heart rate, etc. Subjective methods used

(SA) was formally defined by Endsley as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [110, p. 36], [111, p. 258]. This construct allowed investigations into the quality of the information being processed by human in the context of the information being displayed and acted upon. It addressed the perception and cognition stages in Figure 2-2. Techniques were developed to measure SA as well.[21] *Trust* was defined by Lee and See as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" [114, p. 54]. Modern user-centered design principles aim to calibrate human operators' trust of non-human system components by focusing on how trust develops, how much computers and automation are relied on, and how trust should be managed over time [115]–[120].

Measuring workload, SA, or trust presents some challenges, akin to the social-science limitations discussed in Section 2.1.1. Various implementation difficulties can exist depending on how detailed the experimentation aims to be. Measuring equipment can be intrusive, it can be difficult to diagnose the specific construct being targeted, skilled operators may not accept the experimental setup, or measurements may be inconsistent or unable to detect changes. Despite these limitations and the difficulty in producing ecological validity from controlled experiments, human studies have yielded a wealth of useful information that has influenced engineering best-practices in many ways. Various design standards and specifications exist for HSI that consider physical and cognitive variability, workspace climate and design, arrangement and formats of controls and displays, display principles, cues and alarms, manual handling qualities, and management of physical work, and sleep and fatigue schedules [28], [40]–[42], [121]–[128].

The user-centered viewpoint tends to focus on the limitations of humans to determine how systems should be designed. It has led to in-depth practices for functional allocation between humans and machines and guidelines for automation design strategies. It has forced designers to ask *who* has the final control authority in a system, and under what context(s). Mode confusion is explicitly targeted by providing humans more knowledge of the software abilities and salient cues during mode changes. Systems are designed to gracefully degrade through explicit stages of Sheridan's levels, with minimal interaction required by the human so as to avoid alarm fatigue and workload limitations. In theory, the human is to be protected from biological limitations, so as to prevent human error from jeopardizing the system [69].

One challenge with marginalizing the human's role in the systems occurs if the full problem-space is underrepresented to the human. Ashby, the leading system theoretician in the 1960s, introduced the *Law of Requisite Variety* [129]. Checkland summarizes it this way: "Effective control in a changing environment requires a controller with a variety of response which can match the variety of the environment" [45, p. 88]. A retreat from the SSSI mentality that plagued the technology-centered viewpoint is warranted. However, by

questionnaires like the Modified Cooper-Harper, the NASA task load index, and the subjective workload assessment technique [106]–[109].

[21] These included the Situation Present Assessment Method and the SA Global Assessment Technique [112], [113].

intentionally limiting—by design—the information at least *available* to the human, the role of the human in the system becomes marginalized. In fact, it has been shown that expert human operators are always naturally building abstract state representations of their problem domains to successfully manage complexity [82], [130].

Another problematic area is that detailed cognition models are impossible to prove or disprove [4]. It is easy to overgeneralize the symptoms of systemic problems by referencing non-observable cognitive mechanisms. There is also a tendency to put the blame in psychological terms by substituting deficiencies in performance with cognitive constructs (e.g., SA) that simply "fail" [131]. Focusing on constructs that can only be measured in a laboratory, indirectly at best, unduly shifts the focus away from investigating system-wide task performance in the real ecology [132].

Brunswick warned in 1956 that "psychology has forgotten that it is a science of organism-environment relationships, and has become a science of the organism" [133, p. 6]. The information-processing (user-centered) paradigm sought to treat the human as the weakest link and simply eliminate the opportunity for human error. Some hoped that eventually there would be no need for HF research [134]. However, advancements in thinking about the human role in complex systems, from a systems-theory standpoint, maintain that while failures in sociotechnical systems always involve a human contribution, the focus for preventing performance degradation should not be on blaming a human (or any single a mechanical component, for that matter) [135]. The approach to designing and regulating complex human-machine systems should consider the humans as beings *within* the system, and engineers should avoid the reductionist techniques of evaluating human performance in isolation from the true system they are meant to be a part of.

Incorporating a modern system-safety approach for product testing requires analysis techniques that appropriately account for human contributions to safety. The next section looks at the historical role of humans in systems from the perspective of safety. It then discusses a modern, system-theoretic view of the human. This leads into the introduction of STAMP, the safety model chosen for study in this thesis.

### 2.1.3 Progress in Safety

Until recently, accidents were considered only the result of component failures or human errors. Safety, with this mindset, could be assured by maintaining reliable components. This put an onus on machine maintainers to prevent accidents via regular preventive measures or quick fixes. They would be blamed if they did not ensure properly functioning parts. The human users would be blamed if they failed to notice or prevent mistakes during operation. In the early to mid-1900s human failure was seen as a consequence of poor operator selection; people needed to have the right natural skills and aptitude to operate and maintain their systems [1], [61]. Once the user-centered viewpoint began to permeate human engineering, the blame on operators was simply shifted into psychological terms

(e.g., the pilot lost SA[22]). Doing this did not change the fact that humans were treated as failing components.

System safety began developing as a movement in the aviation community during the 1940s, motivated by the desire to formalize proactive design approaches for accident prevention. The *fly-fix-fly* approach of correcting individual design problems was inefficient and deadly. Although it was effective in reducing the repetition of mishaps with identical causes in systems whose designs and technologies evolved slowly, it was not appropriate in new systems incorporating the latest technology, whose accidents were too costly to use as design lessons [136]. C. O. Miller, one of the leading proponents for establishing system safety over the following two decades, cited a paper by Amos Wood of Boeing that called for improvements to safety practice, including the continual focus of safety during system design, statistical databases for accident analyses with a sharing of lessons learned, and safety education [85].

The military was particularly invested in this new movement, as its recent direction toward developing ICBMs would result in machines that would not have a human operator if problems arose once airborne. MIL-STD-882 (System Safety) was developed to complement the system lifecycle approach of the newly emerging systems-engineering movement. It prescribed safety management not only during design, but during development, procurement, and operation. MIL-STD-882 is still the primary reference document for system safety, not only in the military, but in many industries as well [48]. It serves two primary functions. First it is a systems-engineering management guide for the safety-assurance tasks that must be accomplished during a system lifecycle. These include identifying hazards, performing hazard analyses, mapping subsystems to top-level requirements, and eliminating or reducing risk through the various stages and sub-stages of system development and acquisition. Second, it establishes the fundamental safety concepts and definitions which influence the various regulations, policies, and practices covering the many facets of system safety.[23]

When it was created, MIL-STD-882 incorporated the generally accepted definition of risk at the time (and still does): "A combination of the severity of the mishap and the probability that the mishap will occur" [11, p. 7]. This definition aimed to quantify danger in order to inform the analytical systems-engineering process. The growing system-safety movement favored a traditional accident model that would allow numerical Probabilistic Risk Assessments (PRA) to be calculated. The focus of this model was on physical-component reliabilities and individual failures that connect with each other in time and/or space to cause accidents. This type of model is evident in chain-of-events hazard analyses like Failure Modes and Effects Analyses (FMEA), Fault Tree Analyses (FTA), and the

[22] Dekker points out that it is silly to think that a person loses situation awareness. People are always aware of the situation as they see it, and they build their world based on the feedback they receive from it [4].

[23] MIL-STD-882 does not in itself contain operational mishap prevention or accident-investigation guidance. These responsibilities are governed by separate policies and practices specific to each industry and its regulators. MIL-STD-882 also does not contain design specifications. In aviation, that responsibility has grown into the modern airworthiness certification processes [28], [137], [138].

"Swiss Cheese" model [139]–[141]. The chain-of-events mentality is human nature. *Hindsight bias* causes humans to find critical points in a story of the past in which to inject common sense on what components of a system should have been behaving appropriately or what decisions an operator should have been making [3]–[5]. It is a feature of what Taleb calls the *narrative fallacy*, which in part drives people to draw an "arrow of relationship" onto a sequence of events [2, p. 64]. When simple relationships do exist—as they did more often during the infancy of the human species before elaborate sociotechnical systems[24] existed—breaking the chain is often enough to stop the arrow.

PRA techniques rely on two methods in order to quantify safety hazards, inspired by work in epidemiology in the 1940s [142]. The first uses analytic reduction to separate physical components in space or time. Failure probabilities can be propagated through all the possible branches of causation that were thought of by the analyst. This method has to assume that components fail with independence, and that the reliabilities of all the components are fixed and known. However, the simple causal relationships modeled in event-chain analyses cannot *predict* which pathways will actually fail; they also overly assume component independence and they are difficult to keep updated as the system evolves [143]. Additionally, the interactions of components and any emergent behaviors that cannot be quantified are omitted from the analysis.

The second method uses statistics to look for patterns and regularities in the behavior of a sociotechnical system, a technique influenced by the biological and social sciences [45]. A specific quantity (e.g., number of crashes per year) is measured consistently and correlated to known or measurable changes in design or use policy. This method treats the system like an unstructured mass with interchangeable parts and invokes the *Law of Large Numbers* to predict expected future results.[25] This relies on *regression to the mean* and requires more historical data than are generally available. Statistics are difficult to apply inferentially without strict controls of how data were produced; thus, this method cannot determine explicit causalities, much less in representative detail. It is also vulnerable to the *Law of Small Numbers*, in which observers make judgements based on either too small an amount of data or data from a sample not representative to the entire population [145]. Simulations can be used to produce large quantities of results, but they are valid only within the assumptions that they are based on.

The behavior of humans and software is usually managed arbitrarily in PRA. Human reliability is quantified through approximations from laboratory studies on human errors, where "researchers test limited, contrived task behavior in spartan settings" that does not necessarily "export to natural settings where people carry out actual complex, dynamic and interactive work" [135, p. 62]. Software is certified to prescribed levels of reliability. However, software is a design of a machine abstracted from its physical realization, and it cannot fail any more than blueprints can fail; often software "errors" are a result of poor behavioral requirements [53].

---

[24] This term was coined in the 1950s to refer to complex work domains influenced by both technology and by human behavior and social infrastructures [6].

[25] The average of the results obtained from a large number of trials should be close to the expected value [144].

The intent of the system-safety movement is to acknowledge the complexity of systems by being proactive in systematically tracing hazards throughout the design and life of a system. However, that is not enough for accident prevention. The traditional accident model oversimplifies hazard analyses. A newer, systemic approach to system safety avoids reductionism, and its determinations do not have to be quantitative to be useful. Cybernetics, as a qualitative system-theoretic discipline, demonstrated this. Slovik, as cited by Kahneman, said [145, p. 141]:

> Risk does not exist "out there," independent of our minds and culture, waiting to be measured. Human beings have invented the concept of risk to help them understand and cope with the dangers and uncertainties of life. Although these dangers are real, there is no such thing as "real risk" or "objective risk."

Although risk might not be objective, system properties can be. Systems theory acknowledges that values and goals exist at the top level of a human-activities system [45]. This allows each level in the hierarchy to set a cost function for the lower level. Any assumptions or uncertainties at any level can be interpreted in the language and meaning of the level above it. Uncertainty might be qualitative, but its traceability would be manageable. A system-theoretic model for risk management in sociotechnical systems was originally proposed by Rasmussen, shown in Figure 2-5 [7]. It represents a fielded system having a hierarchical control structure including all levels of society and policy down through the organization performing the operations. Safety practice could be accomplished by a systems approach that gives stakeholders a functional view of their organization and operations, allowing behavior constraints and work boundaries to be implemented and enforced.

Kahneman said, "Policy is ultimately about people, what they want and what is best for them" [145, p. 141]. Design of both hardware and software extends policy to govern components, relationships, and communications. In complex system behavior, the things that matter to stakeholders, designers, and users are not usually the individual physical components; instead what matters are the emergent properties (states and behaviors) that come from the interactions of the components. A modern accident model should be able to acknowledge control, communication, and emergent behavior to reveal the causal *scenarios* underlying desirable (and undesirable) behavior.

The traditional accident model of decoupled components and linear causation—along with the technology-centered and user-centered viewpoints of human-machine systems—treats humans as the last defense before an accident. However, safety, or the lack thereof (accidents), is an emergent property of a human-activity system, and it only has meaning and relevance at the level of stakeholders of the system. Reliable components do not solely assure safe operation, nor do component failures or human errors guarantee accidents. Furthermore, failures alone are too simple an explanation when accidents do happen.

A comparison between the traditional and the system-theoretic safety views is reproduced from Leveson in Table 2-1 [5]. The shift away from assigning blame to humans does not marginalize their roles in the system, but rather makes them a crucial part of it.

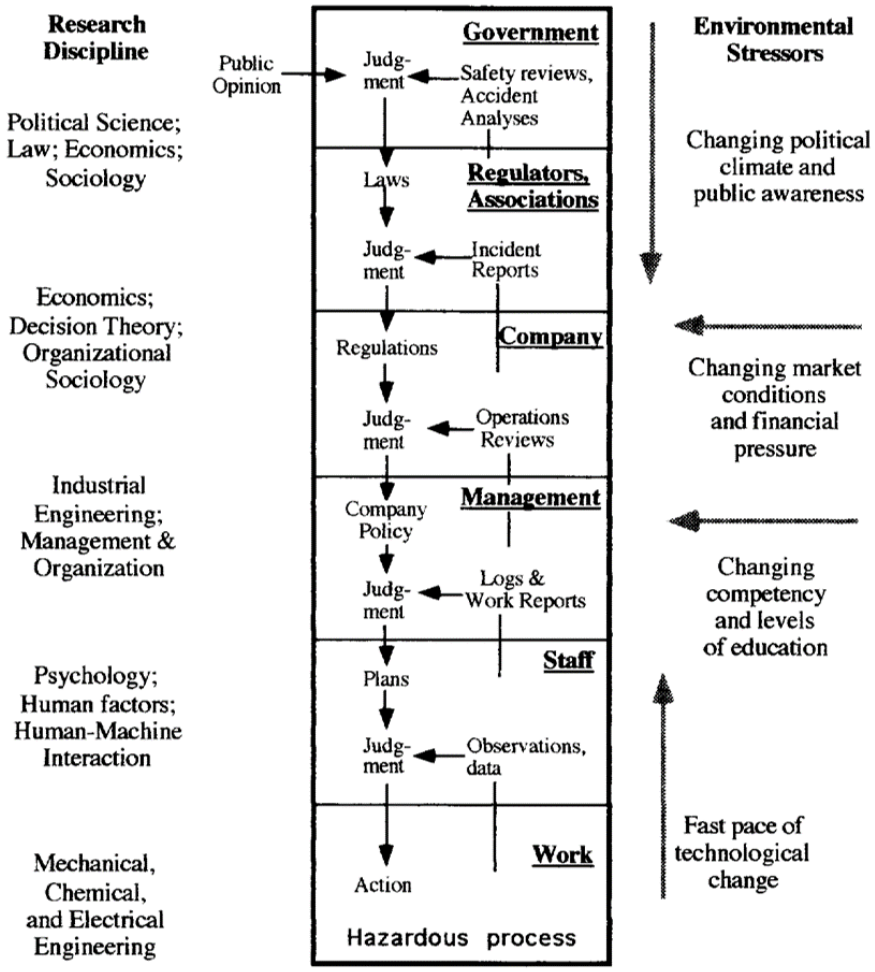



Figure 2-5. Fielded-System Control Structure [7]

### 2.1.3.1 System Theoretic Viewpoint of the Human

The user-centered viewpoint of HF was underpinned by behaviorism and cognitivism in psychology. The information-processing cognition models that dominated the latter half of the twentieth century treated people as machines with components of thought processes and memory, each with mental capacity limits. Normative models of decision-making were assumed, and if humans violated logic it was because of mental biases and process limitations [3], [145]. However, that viewpoint's emphasis on preventing humans as a source of *error* was ignoring the strengths that humans bring to systems, namely intuitive, flexible decision-making and improvisation. Humans demonstrate their strengths best when there are "potential disturbances, uncertainty about the process, uncertainty about the system environment, or all three" [21, p. 93]. Additionally, because humans determine the top-level values of a system, they are also the source of *safety*, both as designers and operators [10], [146].

Table 2-1. Updated Views of Safety [5]

| Traditional | Updated |
|---|---|
| Safety is increased by increasing system or component reliability; if components do not fail, then accidents will not occur. | High reliability is neither necessary nor sufficient for system safety. |
| Accidents are caused by chains of directly related events. | Accidents are complex processes involving the entire sociotechnical system. Traditional chain models can't describe this process adequately. |
| Complex accidents occur from chance simultaneous occurrence of random events | Systems will tend to migrate toward states of higher risk. Such migration is predictable and can be prevented by appropriate system design or detected during operations. |
| Probabilistic risk analysis based on event chains is the best way to assess and communicate safety and hazard information. | Risk and safety may be best understood and communicated in ways other than probabilistic risk analysis. |
| Highly reliable software is safe. | Highly reliable software is not necessarily safe. Increasing software reliability will have only minimal impact on safety. |
| Most accidents are caused by operator error. Rewarding safe and punishing unsafe behavior will eliminate or reduce accidents significantly. | Operator behavior is a product of the environment in which it occurs. To reduce operator "error" we must change the environment. |
| Assigning blame is necessary to learn and prevent accidents or incidents. | Blame is the enemy of safety. Focus should be on understanding how the system behavior as a whole contributed to the loss and not on who or what to blame for it. |

A system-theoretic[26] viewpoint of humans began to evolve in the 1980s motivated by *constructivism* in psychology [93], [147]. Influenced by the work of Piaget and Bartlett, the constructivist perspective does not treat the world as being external to the human (objectivism), but instead views it as a structure that is mapped within the mind, where meaning is created from experience [148]. Mental representations of an associated set of perceptions, ideas, and/or actions are organized into a *schema*, and humans are constantly trying to reach equilibrium between the work domain and their internal construction of it.[27] The human mental model is not static but rather reflects the integration of experience over time, and making errors is actually beneficial for learning [150]. Moreover, the human-activity system (work domain) is itself changing, so these adaptive internal models are necessary to prevent humans from becoming brittle controllers [49].

[26] It can also be called "use-centered" or "ecological" [62].

[27] This is accomplished by the complementary mental activities of assimilation (perceiving new objects or events in terms of the existing schema) and accommodation (restructuring the schema to provide consistency with external reality). Piaget developed these concepts into his theory of child cognitive development, which was later adapted by Perry to describe adult learning [149].

Schemata (within the mental model) guide expectations as well as actions while simultaneously being shaped by consequent experiences. Humans are controllers within the system (not outside it), whether they are the pilot of an airplane or the manager of a company, and the communications and feedback to them can be modeled. How these channels of information govern their actions can be examined, but more importantly, these analyses are only really valid in the natural work ecology.[28] While the technology-centered and user-centered viewpoints provide useful insights into the nature of human work, the system-theoretic viewpoint argues for a departure from laboratory paradigms in order to understand and manage "the context-sensitive nature of performance" [62, p. 298]. Checkland puts this in terms of cybernetics when he paraphrases Bateson [45, p. 86]:

> The mind-body problem which has dogged philosophy for hundreds of years, viewed cybernetically, seems to be wrongly posed. In a cybernetic analysis of the process in which a person thinks, acts, and modifies subsequent behavior in the light of preceding acts, all these items (including the acts themselves) may be seen as information processing. The total self-correcting unit that does this processing is not, however, the human being; it is a system whose boundaries extend beyond the human body. The system is a network of information—transmitting pathways including some external to the actor; on this view, mind is not simply associated with the human body but is imminent in the brain, plus body, plus environment.

The idea that meaning is created within the full ecology—encapsulating both the material phenomena and the experience by the mind—reflects a non-dualistic ontology that can inform engineering problems [133], [151]–[153]. The meanings of interactions between human awareness and work situations exist in terms of the functional significance to the entire system, which makes meaning an emergent property [69]. Gibson hypothesized that a human uses *direct perception* to structurally map available representations of the world and their associated meanings; any particular human naturally sees *affordances*, which are the opportunities for (and consequences of) his actions in a particular ecology [154]. Suchman and Hutchins developed the concept of situated cognition, which treats human actions as purposeful within a particular situation [155], [156]. Furthermore, different controllers (human and non-human) can share the cognition required to manage complex processes, and their joint behavior satisfies meaningful objectives at a higher level of the system hierarchy [157]–[160].

Human decision-making theories also evolved with the shift to system-theoretic views of human engineering. Normative models had become very elaborate, most notably in work by Gilovich and Kahneman that identified mental heuristics—biologically evolved biases[29]—that ease cognitive load in humans [3], [145]. Descriptive decision-making

---

[28] The system-theoretic viewpoint should not be seen as a replacement for the user-centered viewpoint, but an evolution. It is still important to analyze human mental processes, but equal consideration should be given to the real-world ecology and how humans adapt to perform work in their environments. Limited research efforts such as laboratory simulations do not focus on anything outside the human. Any systems approach should treat the humans equally with other parts of the system.

[29] For example, the *representativeness* bias drives people to use stereotypical characteristics that they have assigned to certain classes of objects in order to identify newly observed objects; the *availability* and

models began to develop that acknowledged the shared meaning between humans and the environment in professional work domains. These include muddling through, organizational sensemaking, naturalistic and recognition-primed decision-making, and Rasmussen's S-R-K (skills, rules, knowledge) framework [161]–[165].

In recognition of the shift to the system-theoretic human viewpoint, Rasmussen sought a convergence of HF, behavioral sciences, decision theory, and management research into a discipline called cognitive systems engineering (CSE) [8]. The S-R-K framework was his explanation of how a cognitive agent achieves meaningful action in the ecology based on different methods of structural mapping. *Skills*-based responses are automatic and triggered by raw *signals*; *rules*-based responses require the agent to recognize *signs* in the representation and to draw metaphors and associations to previously successful task accomplishment; *knowledge*-based responses require logical *symbolic* interpretation of the data and systematic projections of possible future states in order to interpret consequences and compare alternatives [166].

Rasmussen also suggested a format for the work domain itself to be represented hierarchically. Using a table, a human-activity system can be broken down first into a part-whole decomposition, in which each level represents a whole component made up of the components that exist in the level below. Second, a means-ends abstraction can be developed, in which each level represents accomplishment that satisfies the meaning of the level above. The *what* in a particular level achieves the *why* of the level above by the *how* of the level below. Figure 2-6 shows an example of the resulting abstraction-decomposition space as updated by Lintern [167], [168]. The horizontal axis is a hierarchy of the syntax (hard system), while the vertical axis is a hierarchy of the semantics (soft system).

Although cost functions can be implemented through the hard-system control hierarchy, its capabilities become brittle when the environment changes and even more so when the system goals change. It is in these situations that human controllers within the system are necessary. Humans can afford multiple possibilities for action, and they understand the soft-system hierarchy (the values and priorities of human-activity systems are created by humans). While MABA-MABA certainly informs a human's limitations to process syntax in his local setting, the real strength of humans lies in their ability to put their actions in the context of the entire system and specific ecology. Similarly, while task-analysis methods of the past identified physical possibilities for action, more modern work-analysis techniques aim to acknowledge the means-ends abstraction [169]–[173].

During fact-finding interviews performed for this research, I learned of an anecdote among former Air Force F-117 pilots.[30] It was broadly accepted by that community that

---

*anchoring* biases can cause a person to assume that a particular valuation or set of conditions is more probable in the world because of recently recalled observations of that value or condition; these biases can lead to the human misestimating the true probabilities of observed events.

[30] The F-117 was the first acknowledged low-observable military airplane. Its mission was to strike high-priority ground targets that were protected by enemy radar and air defenses. It would do this by flying intricate profiles that took advantage of the angles in which various types of enemy radar could not achieve detection because of the geometry of the airplane.

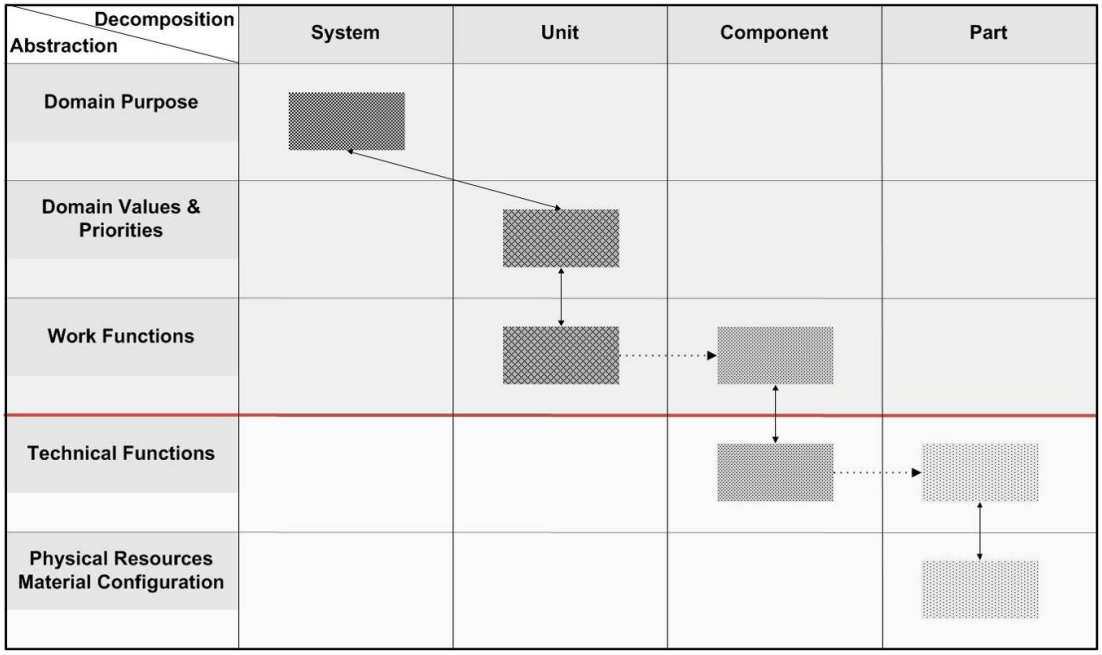



Figure 2-6. Abstraction-Decomposition Space [167]

the autopilot was much better at executing the detailed flight plan than a human pilot. However, the human still had to consent to release weapons when the flight profile brought the aircraft to bear over the target. The pilots would joke that they could have wrapped the consent switch[31] with duct tape because the airplane flew the profile so well, thus negating the need for a human on board. The real theme of the jest was that although it was very capable the autopilot had no means to react to mission-priority changes, unexpected threats, or in-flight problems with the airplane itself. This would have made it a brittle system without a human on board to provide adaptivity.

To be able to adapt while interpreting for the meaningful features of the functional system, humans adopt multi-modal, tiered communications. Steinberg draws the analogies of a human teaming with various partners (e.g., falcon, working dog, ape, and another human) to show the graduation of various features of the communications. One important feature that improves along this spectrum, besides core task competencies, is the ability of more advanced partners to understand each other's actions within the context of the purpose of the work, without the need for explicit communication. Humans not only pick up implicit and non-verbal cues, but the cues they receive can inform their assumptions about higher-level system properties, such as meta-processes and goals [119].

Borst cites Clark on the importance of establishing a "common ground" between able controllers in a system [68, p. 8] before and during operation. The knowledge that a controller has about other controllers' understandings of the levels of meaning in a particular ecology will naturally determine the balance between trust and verification between controllers. If controllers are sufficiently grounded, their communication becomes

[31] "Pickle" button.

more subtle. Feedback to a higher controller might not have to be detailed, but rather contain only a confirmation from the lower controller that it is functioning properly. However, as Ashby warns, when the variety of system capabilities increases, the available variety of communicated information must increase, not decrease [129], [174].

Merlin states that accidents "do not occur because of a single event, but rather from a series of events and actions involving equipment malfunctions and/or human factors" [175, p. ix]. This mindset goes beyond the technology-centered view of naturally-skilled operators that are expected to always break a mishap chain. It begins to assume the user-centered view of well trained, highly knowledgeable and experienced operators using machines designed to support human capabilities—a concept embodied by the Wright brothers ahead of their time. However, this view is still supported by safety philosophies that either ignore humans or treat them as the last line of defense (with some probability of failure). This has resulted in several approaches for quantifying human-error rates, refined from the work of Fitts, Norman, and Reason. Experimentally-derived statistics and taxonomies are used to categorize types of error and inform hard-systems task analyses and designs [141], [176]–[179]. Although human-engineering research has made substantial progress since the middle of the last century, this assumption about safety still forms the basis of the current human-error analysis practice in use by the military, which is discussed in Chapter 4.

Dekker asserts that finding specific events that lead to accidents is no more useful than trying to find specific events that lead to safety, and that classifying errors is not the same as explaining how hazards develop [4]. Blaming a human for the cause of an accident (or for a singular act that saves a system from an accident) is reductionist in scope: it views a human as an outside observer and a final point of a failure chain. A system-theoretic approach instead acknowledges that human-activity systems are dynamic and serve a purpose, and that humans introduce variability *throughout the life of the system* that is necessary for safety [131]. Humans exist within the system, and they contribute to safety and other top-level system properties because they are goal-seeking decision makers. Looking at systems along the vertical axis of the abstraction-decomposition space calls for the consideration of what types of controls, communications, and feedback are required at each functional level of the system to ensure its meaning and purpose are understood and preserved.

### 2.1.4 STAMP

The traditional view of safety considers physical-component reliabilities (including people and software) and identifies failures that propagate to accidents. In the middle of the last century this view gave rise to FTA, FMEA and other PRA techniques that are still in use today. A more useful, system-theoretic view of safety began to emerge near the turn of the century with Rasmussen's work on risk management in sociotechnical systems [7]. Most recently, Leveson proposed a theoretical accident-causality model called STAMP that treats the prevention of undesirable losses as a control problem. Accidents are more than simply chains of component failures; they involve complex, dynamic processes. Safety is considered an emergent property that arises when system components interact with each

other within a larger environment. Hazards are not simply equated to failures, but to states or conditions of the system that can be prevented or mitigated [5].

In STAMP, the goal of safety engineering is to control the dynamic behavior of components—regardless of whether they are human, software, or machine—to ensure that safety is enforced. This modern accident model looks at more than just component failures and includes factors like system design errors, software requirements flaws, human decision-making, and the migration of the system over its lifetime to states of higher risk. It extends the classic accident model while still including failed components as a subset of possible causal scenarios, and it is able to capture component interactions and emergent behaviors which are typically omitted from traditional hazard analyses.

Accident investigations based on STAMP use a technique called Casual Analysis using Systems Theory (CAST). CAST identifies the behaviors and scenarios which contributed to a mishap within a sociotechnical system. Relationships between all associated behaviors and decisions are documented to minimize emphasis on any one component or actor and reduce hindsight bias. STAMP is also used for mishap prevention—even when there may not be any past incidents from which to draw experience—in an hazard-analysis technique called System-Theoretic Process Analysis (STPA). STPA is an engineering approach that identifies explicit causal scenarios for accidents.[32]

Systems-engineering management must typically document a list of possible accidents, the modeled system and its components, and hazard scenarios that could cause the accidents [11]. Although a hazard analysis based on the traditional accident model fits the criteria demanded by the systems-engineering process, it reduces accident prediction to an outcome of independent component integrities. FMEA, for example, is *bottom-up*, starting with the physical system as designed and propagating component failure probabilities. It fosters a tendency to apply safety mitigations after the design, which demotes safety from being a top-level and early consideration of the system. STPA is *top-down* by nature.[33] It complements the systems-engineering tenet of hierarchical decomposition and establishes a model for hazard analyses that can be maintained by system managers. In this way safety hazards can be traced and evaluated throughout all stages of the system lifecycle, including the conceptual stage [185]. If STPA is used early in the lifecycle, the results of the analysis can be used to generate system and subsystem requirements and create a safer design from the start.

The three basic concepts in STAMP are: a) safety constraints, b) hierarchical control structures, and c) process models [5]. The *safety constraints* derive from identified accidents and hazards. The *hierarchical control structure* is the representation of the functional system and its component relationships, allowing the actions and processes between components to be examined in light of the safety constraints. The *process model* of each controller is its understanding of the states of the process it is controlling. The

[32] For STPA examples, see Antoine [180]; Dunn [181]; for CAST, see Hickey [182]; Nelson [183].
[33] For a good comparison between a PRA approach and STPA, see the Leveson at al. analysis of civilian airworthiness guidance [184].

following four sections briefly discuss how STPA is conducted. Sections 2.1.4.1 and 2.1.4.2 discuss the identification of accidents and hazards and the safety (hierarchical) control structure. Those steps are necessary for the subsequent application of STPA and should be accomplished naturally through the implementation of appropriate systems-management processes. Sections 2.1.4.3 and 2.1.4.4 discuss steps specific to STPA.

#### 2.1.4.1 Identifying Accidents and Hazards

The following definitions are reproduced from Leveson and are based on systems-theory and system-safety concepts [5]:

*Accident* – An undesired or unplanned event that results in a loss, including a loss of human life or human injury, property damage, environmental pollution, mission loss, etc.

*Safety* – Freedom from accidents.

*System Boundary* – Domain within which the system designer has control, outside of which is considered the environment.[34]

*Hazard* – A system state or set of conditions that, together with a particular set of environmental conditions, will lead to an accident (loss).

Only organizational stakeholders can identify accidents, because accidents equate to a loss for the entire enterprise or culture that created or inherited the human-activity system. Because STAMP is a top-down model, a necessary precursor to a hazard analysis is a well-defined set of accidents. Safety investigators can guide the discussion to define the accidents, but the stakeholders own the process. Accidents may also be prioritized by severity, because the prevention of *all* accidents is never a 100 percent guarantee (the only perfectly safe system is the one that is never built or operated). The list of accidents need not be extensive.

After accidents, hazards are identified. The list of hazards is typically a small, exhaustive set because hazards do not include any detailed engineering assumptions from the design (e.g., “toxin is released into the environment” is a more appropriate hazard than “valve leaks and releases toxin”). Each hazard will trace back to one or more accidents. The prevention of a hazard is stated as a safety constraint. Losses can be mitigated only by making efforts to enforce safety constraints, ensuring the system does not enter a hazardous condition.[35] The hazards (and thus safety constraints) should be revisited during safety planning and throughout the life of the system.

---

[34] In open systems (including human-activity systems) there is a constant exchange of materials, energy, and information between the system and the environment. Designers may enforce controls and constraints on the system, not the environment.

[35] The 2011 Fukushima nuclear power plant disaster was a combination of environment (in the form of a 15 meter tsunami caused by an unanticipated earthquake) and a hazard (emergency diesel generators were located in the basement of the turbine buildings). The generators were unfortunately in a location that was easy to flood, and the sea wall was only built to stop wave heights predicted by standard tsunami models. The location of the generators was under the control of designers. The strength of the tsunami was not [186].

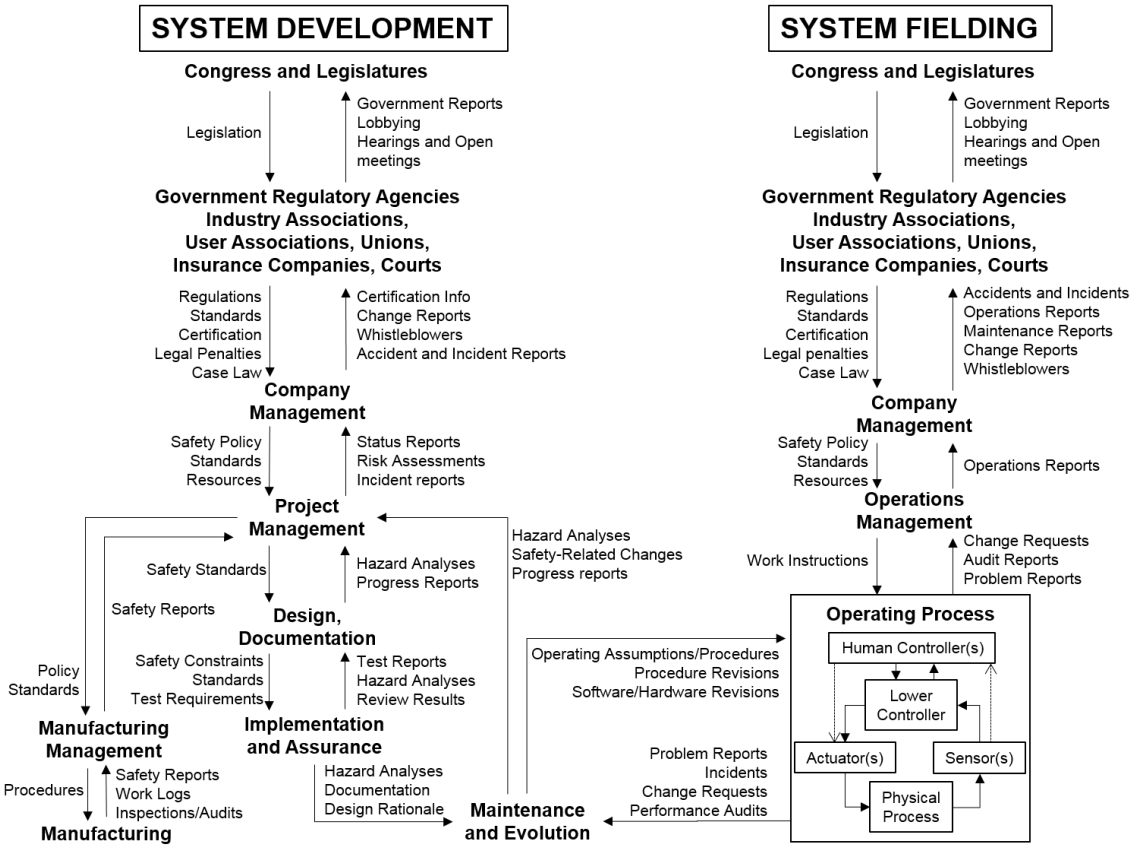



Figure 2-7. Organizational Control Structure Example [5]

The hazards than can lead to accidents should not be confused with the losses themselves. An example is the Navy's SUBSAFE program, instated after the 1963 USS Thresher incident involving a deep water flood following engine room pipe leaks [187]. Organization-wide, the only two safety constraints of SUBSAFE are: maintain hull watertight integrity and maintain operability and integrity of critical systems to control and recover from a flooding event. The existence of a flooding event does not necessarily mean that the submarine has completely flooded or that there has been a loss.

### 2.1.4.2 Safety Control Structure

The hierarchical safety-control structure is an abstract model of the system that illustrates the relationships between functional levels of control (whether they be people, computers, components, organizations, entities, etc.). These relationships are based on systems theory, and each level imposes constraints on the behavior of the lower level. Leveson produced a generalized example of a STAMP organizational control structure to extend Rasmussen's operational risk management diagram from his work in CSE, reproduced in Figure 2-7. It is a decomposition, starting at the top with legal/regulatory and organizational entities all the way to the lower-level components of the system operations. STAMP treats safety as a combined problem between systems-engineering development and field-use. Stakeholder

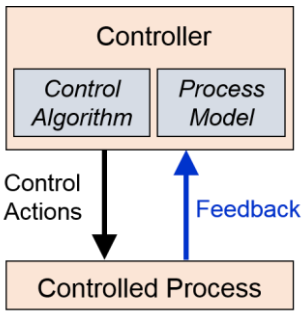



Figure 2-8. Basic Control Loop

and designer requirements and constraints are communicated horizontally between these two major stages of a product lifecycle.

The model is not necessarily a representation of the physical structure of the system. For example, suppose an air-traffic controller speaks to a ground-based remote operator of a flying drone. This communication occurs through a radio signal that travels from the control tower to the drone, is multiplexed and sent to the operator's ground control station via datalink transmission, and then demultiplexed into an audio channel in the operator's headset. While a physical schematic would detail the intricate connections just described, the safety-control structure would show the tower personnel functionally controlling the drone operator, who in turn would be functionally controlling the drone.

The purpose of the safety-control structure is to enforce safety constraints and therefore eliminate or reduce losses. A complete control structure contains not only information exchanges, but command relations,[36] control roles, and safety responsibilities. Control loops exist between every level of the safety-control structure, including the management and organizational levels. Loops between lower levels typically operate with a shorter time constant than those between higher levels [185]. Figure 2-8 shows a very simple feedback control loop. The controller is assigned safety constraints to enforce on the controlled-process behavior, which it does by issuing control actions to change the state of the controlled process. Every controller contains an algorithm for deciding what control actions to provide. A controller also holds a process model (or mental model in humans) of the current state of the process being controlled.

An advantage for system managers is that by defining and periodically refining the hierarchical control structure, it can be used as a shared visual-planning tool among stakeholders [56]. The proactive use of an organizational model is described by Stringfellow as exhibiting "mindfulness or heedful interrelating…important for a resilient organization" [21, p. 94]. The model includes many types of relationships, including official, secondary, backup, and sometimes unofficial controls and communications. This makes the control structure much more than an organization chart or a design schematic. However, the design documentation, as well as the sources shown below (reprinted from

[36] Higher levels in the hierarchy have broader responsibility, authority, and accountability than lower levels.

Stringfellow), can be used to help put together the complete control structure to include the many levels of the sociotechnical system [21]:

- Organization Charts
- Documented Activities
- Problem identification and resolution processes
- Reports
- Manuals, Policies, and Procedures (including emergency procedures)
- Funding sources and channels
- Hiring Policies

Additionally, more rigorous model inclusion criteria have been developed by Dulac and Stringfellow, and those are discussed and expanded in Chapter 4 [21], [188].

#### 2.1.4.3 STPA Step 1: Inappropriate System Behavior

Because STAMP treats safety as a control problem, STPA frames its analysis around inappropriate control actions that can lead to hazards. These are called unsafe control actions (UCA), and there are four general types:

A) A necessary control action is not provided to avoid a hazard.

B) A control action is provided that creates a hazard.

C) An appropriate control action is provided, but it is too early, too late, or in the wrong sequence.

D) An appropriate control signal is sustained too long, stopped too soon, or at an inappropriate intensity.

Each control loop must be examined to identify UCAs. For every control action that a controller is capable of issuing, each way in which it could be unsafe is documented by also specifying the *context* in which it is unsafe. Additionally, the top-level hazard(s) the UCA would lead to is/are noted. The context is important because, in any reasonably designed system, there is only a subset of conditions under which the control actions are unsafe, and one of the purposes of Step 1 is to identify them. If the control action were always unsafe, it would not be included in the design. The UCAs may be documented in any format, but the default is to use tables that help to organize the analysis by referencing the four types of UCA. Table 2-2 shows an example template (note that there can be multiple or no entries in each box).

Once the UCAs are identified, the second step of STPA is to identify the potential causal scenarios contributing to unsafe control. It can be convenient to separate STPA into two steps (as is done here) by first identifying all the UCAs and then the scenarios that can cause the UCAs. However, this is not necessary, and the two steps could be combined in some other fashion, such as identifying each UCA and immediately looking for its causes.

Table 2-2. Example of a Step-1 Template

| Control Action | Unsafe Control Actions | | | |
|---|---|---|---|---|
| | Not Providing Causes Hazard | Providing Causes Hazard | Wrong Timing/Sequence Causes Hazard | Wrong Duration/Intensity Causes Hazard |
| A | "A" not provided<br>when/while *context(s)*<br>[List potential hazard(s)] | "A" is provided<br>when/while *context(s)*<br>[List potential hazard(s)] | "A" provided too early/late<br>when/while *context(s)*<br>[List potential hazard(s)] | "A" applied too short/long<br>when/while *context(s)*<br>[List potential hazard(s)] |
| B | "B" not provided<br>when/while *context(s)*<br>[List potential hazard(s)] | "B" is provided<br>when/while *context(s)*<br>[List potential hazard(s)] | "B" provided too early/late<br>when/while *context(s)*<br>[List potential hazard(s)] | "A" applied too short/long<br>when/while *context(s)*<br>[List potential hazard(s)] |
| etc. | | | | |

Additionally, a master table can be maintained (not shown) that documents the traceability from UCA's to hazards and from hazards to accidents.

#### *2.1.4.4 STPA Step 2: Causal Scenarios*

STPA uses UCAs to guide the analysis for potential scenarios for hazardous behavior, including actions by software or humans. Causal scenarios explain accidents more explicitly than simply stating the controller failed with no further information about why. Without understanding the causes of unsafe behaviors, options for eliminating or reducing them are limited. Figure 2-9 is an expanded view of the control loop in Figure 2-8 that shows some scenarios that could contribute to unsafe behavior [5].

Step 2 of the STPA analysis identifies scenarios that can cause the controller to issue a UCA. One way a hazard can occur is that the controller's process model becomes inconsistent with the real state of the controlled process, and the controller provides inappropriate control to the process. Part of the challenge in designing an effective safety-control structure is to provide the feedback and inputs necessary to keep the controllers' models consistent with the actual state of their controlled processes and with each other.

A second set of scenarios identified by STPA involve control actions required to enforce a safety constraint that are provided correctly but not executed. These scenarios involve inadequate behavior (perhaps a failure or a delay) in a part of the control loop besides the controller and its feedback, such as in the actuator or the controlled process itself. If there are multiple controllers providing control instructions to the same process, hazards can result when conflicting control actions are provided, perhaps due to inconsistencies between the individual controller's process models.

The identified scenarios (hazard causes) in Step 2 can be used to eliminate the causes from the system or mitigate them if elimination is not possible or practical.

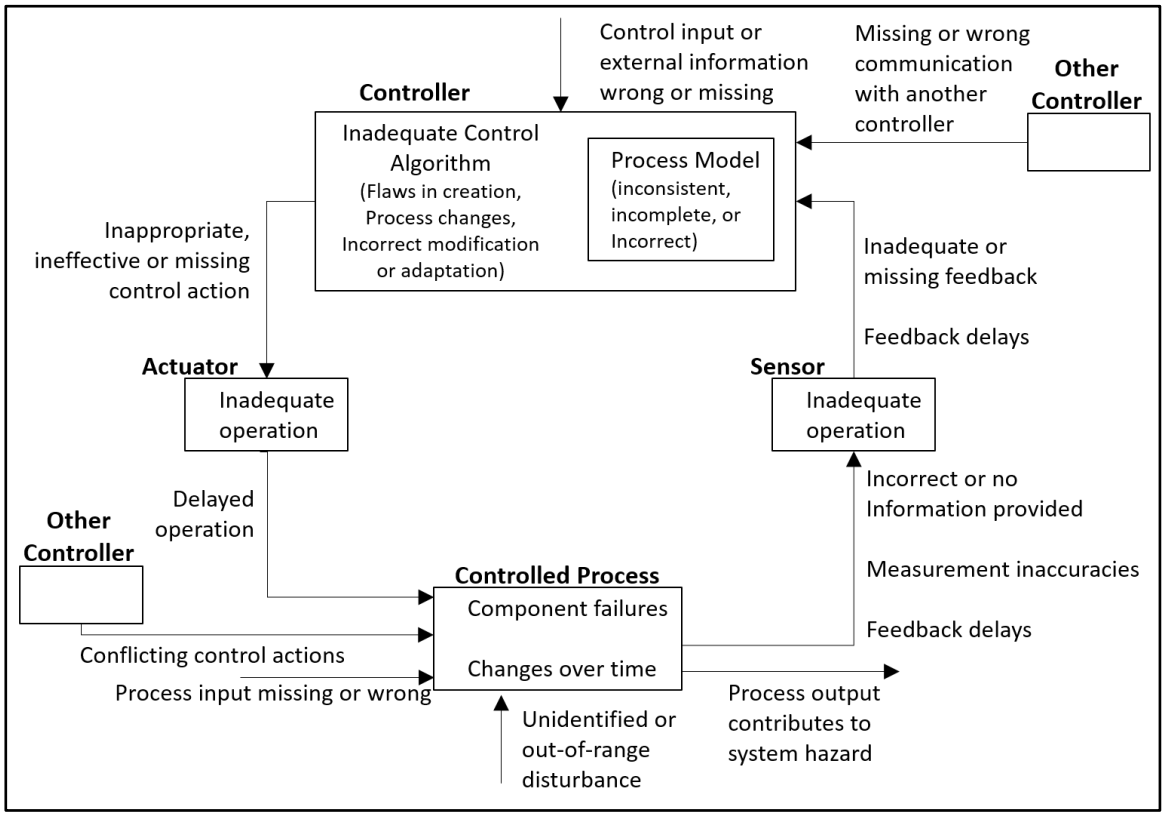



Figure 2-9. Detailed Control Loop [5]

Mitigation might involve changing the hierarchical control structure, parts of individual control loops,[37] or the controller decision-making algorithms.

Finding explicit causal scenarios is a large advantage to a system-theoretic model. There is rarely a single root cause of any modern accident. Step 2 incorporates knowledge about the design, environment, operators, and organizational culture to build scenarios that lead to unsafe control. This requires significant input from domain subject matter experts (SME). STPA gives discipline specialists like software and human engineers a framework to apply their knowledge within the context of the actual design with its safety constraints already defined. This is more powerful than merely applying best practices broadly without regard for explicit safety requirements.

#### 2.1.4.5 Moving STPA Forward

Because STPA is based on hierarchical control, it applies from the organizational structure all the way down to the technical operations. It builds a framework within which experts may guide their discipline-specific experience to analyze the design and intended use philosophy of the system. Like all techniques that use STAMP, STPA maintains a visual

[37] Assigned responsibilities, relationships to the controlled process, allowable control actions, and designed feedback and communications, for example.

engineering model which is beneficial for guiding technical planning discussions among safety planners and stakeholders of different backgrounds.

Safety-guided design (and design-guided safety planning) is extremely useful early in a product lifecycle as it is very expensive to reengineer a system if flaws are not found early. STPA is often used on existing designs or operational systems, because not only can it influence and shape early design decisions, but the analysis can be iterated and refined as the design evolves. Additionally, the ability to formally document scenarios that directly relate to accidents, without relying on the availability of probabilistic estimates, makes STPA useful for system certification [189].

Rasmussen urged that system designers and operators should cooperate in decision-making, with the designers communicating their intent to the operators, and the human operators completing the design [10]. Leveson has advocated methods for specifying, documenting, and updating designer intent [9]. Another important type of information about the system that should be documented is assumptions about the design of the system and how it is used. The current generalized example of a STAMP organizational control structure (Figure 2-7) acknowledges the horizontal transmission of information between designers and users, but only with low-fidelity communication channels containing product maintenance and evolution information.

Product testing is part of "System Development" in that diagram. As discussed in the next section, testing brings with it a host of unique safety considerations and tasks that must be accomplished to steward the design and any assumptions about its fielding. Testing is a method of verifying particular behaviors of the design and documenting the procedures field users will be assumed to use, in order to make modifications to the product and procedures as needed. The existing generalization would benefit from acknowledging a dedicated test stage between development and fielding. By acknowledging a test stage, information about design and use assumptions can be appropriately highlighted and maintained between stakeholders at the different organizations that contribute to the product development and operation. The benefit of STPA during system testing is the focus of Chapter 4.

A limitation of STPA in particular is that it currently oversimplifies the role of humans in systems. Stringfellow began to look at human, social, and organizational factors that contributed to hazardous causal scenarios [21]. Thornberry took the approach of studying the visually-guided Step-2 analysis of the controller to recommend areas for further guidance specific to humans [190]. A diagram like Figure 2-9 is helpful, but it does not differentiate between machine and human controllers. It guides analysts to look at feedback, communications, process model, and control algorithm; this makes it a complete analysis from a system-theoretic standpoint, but it is not refined for humans. This puts a larger responsibility than necessary on human-engineering SMEs to build causal scenarios. The next chapter discusses considerations for how to approach the analysis of human controllers.

## 2.2 Air Force Systems

The motivations introduced in the previous chapter regarding unmanned vehicles, autonomy, and testing merit their own discussion here. Subsequent chapters incorporate background on system safety, human engineering, and autonomy research when new methods are proposed for analyzing human controllers and for performing test-safety planning.

The U.S. Air Force (AF) acquisition enterprise was selected for research application due to its deep investments in developing technologies in software, cybersecurity, and autonomy, as well as their professional product-testing enterprise. All the military services contain similar mechanisms responsible for developing and sustaining new technologies and capabilities, including testing. These practices are guided by the Department of Defense (DOD) acquisition policies and headquarters staff [138], [191]–[199]. The ideals encompassed by the product development establishments within the DOD reflect the development strategies of most prolific, mature product-developing organizations. Those that produce many different types of systems with varying degrees of complexity typically devote an entire sub-organization to testing.

The AF is divided into ten major commands (MAJCOM), which report to AF headquarters in the Pentagon. These major organizations each have a particular portion of the service mission. Each MAJCOM might fulfill one of several capacities: advanced skills training for specific types of operators, development of product requirements and corresponding tactics, techniques, and procedures (TTPs), providing personnel and equipment to a region for mission execution, and creating capabilities that support the mission. Nine of the MAJCOMs are described for reference in Table 2-3 (the tenth, AF Reserve Command, is not shown).

An AF *wing* is an organization that provides an ongoing, self-supported mission under a MAJCOM. Each wing is generally tied to a particular operating location, usually an air base. A wing's singular mission might be combat, training, testing, facilities, etc. Most MAJCOMs have different types of wings. For example, Air Education and Training Command will contain mostly training wings and some others, while Pacific Air Forces will contain mostly combat wings and some others. Between the MAJCOM and wing levels, there can exist an intermediate organizational tier—usually called a *numbered air force* or a *center*—which provides an additional split in the command hierarchy to divide multiple wings into pockets of fewer wings.

Under each wing is a *group*. A group is an organization that provides a specialty to a wing. These specialties include operations, maintenance, medical, personnel management, etc. Even if there is only one wing at a given base, there will be several groups under that wing. Every group contains multiple *squadrons*, which are the execution entities of the group. Within a group these units are identical or similar to each other, and they share the group mission. Finally, some AF organizations do not fall under a MAJCOM but report directly to AF headquarters, either because their contribution is equally applicable to all MAJCOMs or because they are mandated by law to be an independent entity. Each of these is typically designated as a field operating agency or a direct-reporting unit. See Figure 2-10.

Table 2-3. Air Force Major Commands (MAJCOM)

| Name | Headquarters | Specialty | Support | Field Use |
|---|---|---|---|---|
| Pacific Air Forces | Hawai`i | Fighters | | Pacific Theater |
| Air Forces in Europe | Germany | Fighters | | European Theater<br>African Theater |
| Air Combat Command | Virginia | Fighters | Advanced Skills Training<br>Requirements, TTPs | Asia Theater<br>American Theaters |
| Global Strike Command | Louisiana | Bombers<br>Ballistic Missiles | Advanced Skills Training<br>Requirements, TTPs | All Theaters |
| Air Mobility Command | Illinois | Transport | Advanced Skills Training<br>Requirements, TTPs | All Theaters |
| Air Force Special Operations Command | Florida | Special Ops | Advanced Skills Training<br>Requirements, TTPs | All Theaters |
| Air Force Space Command | Colorado | Space Launch<br>Satellite Ops<br>Cybersecurity | Product Development<br>Advanced Skills Training<br>Requirements, TTPs | All Theaters |
| Air Force Materiel Command | Ohio | Technology<br>Product Management | Research<br>Product Development<br>Depot Maintenance | |
| Air Education and Training Command | Texas | Education<br>Training | Military Doctrine<br>Basic Operator Skills<br>Graduate Academics | |

Air Force Materiel Command (AFMC)[38] is the MAJCOM responsible for the development of AF systems. AFMC does not operate systems in the field, but it supports the MAJCOMs that do. The MAJCOMS that do perform field missions (the first seven listed in Table 2-3) are the *using commands*. Under AFMC are four major centers that oversee the primary functions of AF acquisition: research, test, sustainment, and total lifecycle management.

Figure 2-11 shows the basic relationship among the main four AFMC centers. The AF Research Laboratory (AFRL) performs basic and applied research to satisfy technology thrusts supporting all AF systems. The AF Test Center (AFTC) is the enterprise that maintains the profession of developmental test and evaluation (abbreviated DT). The AF Sustainment Center supports systems after they transition to the using commands; it sources parts, performs depot maintenance, and manages in-place modifications in the field. The AF Lifecycle Management Center[39] provides program management and systems engineering, including the research, testing, and field sustainment strategies; it has a large influence over the capabilities and tasks of the other three centers.

The defense acquisition system is intricately tied to a DOD-managed framework called the Joint Capabilities Integration and Development System [200]. This enables users and developers to cooperate in the acquisitions process as early as concept development.

[38] http://www.afmc.af.mil/
[39] Also referred to in this thesis as AFMC program office(s)

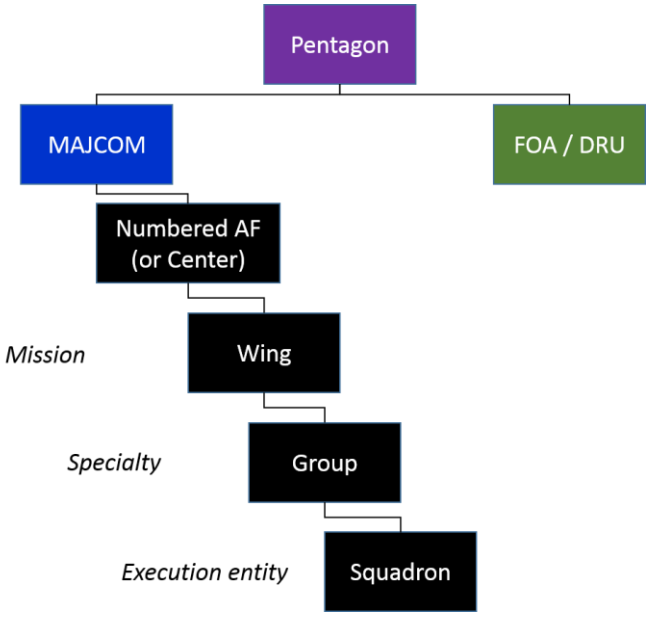



Figure 2-10. Air Force Levels of Command

The AFMC research stage dovetails with the using commands' concept and requirements generation activities. In this stage, technology investigations are planned, and current technological capabilities are assessed against requirements. Capabilities are developed to a technology readiness level (TRL) of at least (6), which means that assembled system components are shown to function outside a laboratory setting. Table 2-4, reproduced from the *Defense Acquisition Guidebook*, gives a reference on how AFMC determines technology maturity [23, p. 848]. Research is supported by extensive modeling and simulation, as well as component bench testing and laboratory experiments.

It is important to delineate between development and fielding. In large industries, after concepts and requirements have been established, products typically go through three stages: research/design, developmental test, and fielding. There can certainly be overlap among these activities, and sometimes items might get fielded quickly with minimal test, but the key principle is that most products go through the formal stages in some manner or another. *Development* covers the first two stages. *Field use* is everything that comes after development. The AF term for field use is "operations," and initial field evaluation is called operational test and evaluation. This thesis instead uses "field evaluations" and "field use" to refer to anything after DT. The term *operations*, or *operating process* in this thesis refer to any active flight process, whether it be during DT or field use.

### 2.2.1 Unmanned Vehicles and Autonomy

Unmanned vehicles (UV) are not the focus of this thesis in the manner of a thorough history or analysis of any particular real-world system. However, UVs provide a good example case for autonomy because of all the developments they have enabled in guidance and control, navigation, and software. Autonomous systems will soon be capable of more than those basic functions, and a hobbyist approach to design and safety planning will not be

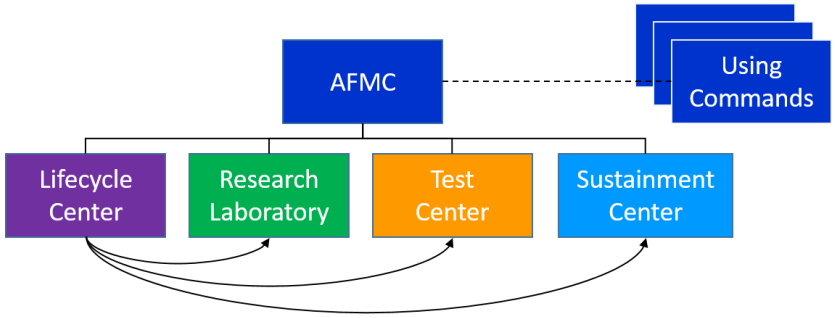



Figure 2-11. Air Force Materiel Command Organization

thorough enough. This section covers a brief history of UVs and discusses the AF research thrusts in autonomy that aim to modernize the design and employment of autonomy-capable systems. In Chapter 4, an experimental UV system provides the use case for the safety-planning method that I introduce for the AF DT enterprise.

#### 2.2.1.1 Evolution of Unmanned Vehicles

Robotic flying systems have been in development for over half a century, beginning with radio-controlled, manually-piloted aircraft and evolving to more complex systems comprised of aircraft that are able to fly autonomously while receiving objectives from human operators in ground stations [175]. While there is no agreed upon term for these systems—a review of the literature yields examples including unmanned/unpiloted aerial vehicle (UAV), unmanned aircraft system (UAS), unmanned underwater vehicle (UUV), remotely piloted vehicle/aircraft (RPV/RPA), remotely operated vehicle/aircraft (ROV/ROA), or simply "drones" [16], [17]—this thesis simplifies the nomenclature by using the term UV. A UV is defined as *a machine that provides conveyance of a passenger or payload while being absent a human operator in or on the machine itself*. The medium in which the UV operates can be made evident by the context of the discussion and qualifying descriptions.[40]

Modern UVs might be manually controlled via radio link or cable; they might also be allowed to perform their own decisions and actions with continuous, periodic, or potentially absent supervision by a human operator. In 2010, the AF began to use the term RPA instead of UAV to emphasize that there are always humans involved in some form in the vehicles' operations. In reality, most modern UV systems operate on a spectrum between manual-remote operation and fully autonomous (or several points along this spectrum). As discussed in the next section, it is in fact difficult to classify any complex system with a broad brush such as "manual" or "autonomous." Automation and autonomy are treated as characteristics of specific functional control loops; complex systems contain many of these loops, and the choice of which loops are important to illustrate and/or analyze is left to the purpose of the task.

UVs provide many benefits compared to their manned vehicle (MV) counterparts for certain applications. A considerable design trade-space becomes free when human

[40] In this broad context, both a pilot-less tram at an airport and a guided ballistic missile qualify as UVs.

Table 2-4. Technology Readiness Levels [23]

| Technology Readiness Level (TRL) | Description |
|---|---|
| 1 - Basic principles observed and reported | Lowest level of technology readiness. Scientific research begins to be translated into applied research and development. Examples might include paper studies of a technology's basic properties. |
| 2 - Technology concept and/or application formulated | Invention begins. Once basic principles are observed, practical applications can be invented. Applications are speculative and there may be no proof or detailed analysis to support the assumptions. Examples are limited to analytic studies. |
| 3 - Analytical and experimental critical function and/or characteristic proof of concept | Active research and development is initiated. This includes analytical studies and laboratory studies to physically validate analytical predictions of separate elements of the technology. Examples include components that are not yet integrated or representative. |
| 4 - Component and/or breadboard validation in laboratory environment | Basic technological components are integrated to establish that they will work together. This is relatively "low fidelity" compared to the eventual system. Examples include integration of "ad hoc" hardware in the laboratory. |
| 5 - Component and/or breadboard validation in relevant environment | Fidelity of breadboard technology increases significantly. The basic technological components are integrated with reasonably realistic supporting elements so it can be tested in a simulated environment. Examples include "high fidelity" laboratory integration of components. |
| 6 - System/subsystem model or prototype demonstration in a relevant environment | Representative model or prototype system, which is well beyond that of TRL 5, is tested in a relevant environment. Represents a major step up in a technology's demonstrated readiness. Examples include testing a prototype in a high-fidelity laboratory environment or in simulated operational environment. |
| 7 - System prototype demonstration in an operational environment | Prototype near, or at, planned operational system. Represents a major step up from TRL 6, requiring demonstration of an actual system prototype in an operational environment such as an aircraft, vehicle, or space. Examples include testing the prototype in a test bed aircraft. |
| 8 - Actual system completed and qualified through test and demonstration | Technology has been proven to work in its final form and under expected conditions. In almost all cases, this TRL represents the end of true system development. Examples include developmental test and evaluation of the system in its intended weapon system to determine if it meets design specifications. |
| 9 - Actual system proven through successful mission operations | Actual application of the technology in its final form and under mission conditions, such as those encountered in operational test and evaluation. Examples include using the system under operational mission conditions. |

operators (and accompanying life support and/or escape equipment) are removed from a vehicle. This can be applied to improvements in performance, efficiency, cost, or payload capacity. Also, with human physiological limits absent from a vehicle, varied environmental opportunities become available. In an aircraft this might translate to higher operating altitudes or more dynamic maneuverability, or in underwater vehicles more severe temperature and pressure thresholds.

UV systems are not without their tradeoffs, however. Currently, it is typical for more human operators to be required for UV operations than in an MV system; they facilitate ground control station (GCS) operations and maintenance. Software design for autonomous operation can be expensive and timely to validate [53]. As autonomy software's state of the art matures, designs are moving toward use scenarios that reduce the ground support human footprint, including having a single human supervisor controlling many UVs [104], [120], [201]. This movement, however, highlights another UV tradeoff: the lack of human awareness of the vehicle's operating environment [202]. There are various visual, aural, tactile, and vestibular channels of feedback that an MV operator

receives from a vehicle and its domain, whereas a UV pilot/supervisor is always divorced from the domain and sometimes experiences communication delays or discrepancies.

With regard to airborne UVs, mechanical component reliability has historically been behind [175]. As these systems evolved—largely in the military beginning in the world-war periods—the lack of risk to on-board pilots coupled with a perceived expendability of the vehicles prompted designers to push the limits of guidance, navigation, and control while ignoring component reliabilities. Hardware has been allowed to eventually become more reliable and less costly, while more focus has been put on software.

When considering mechanical failure in flying machines, it is easy to imagine an end result of a vehicle impacting the ground and harming people or property. This is indeed one type of accident that must be mitigated as airborne autonomy proliferates. The Federal Aviation Administration (FAA) is making efforts to integrate the civil national airspace (NAS) for growing commercial and private use of UVs [20]. Avoiding damage to the general populace, who exist under the NAS, is imperative. The military, while operating in a war zone, might consider ground-impact incidents undesirable for different reasons. For example, some UVs contain classified data or equipment that should not be lost behind enemy lines. However, when training or testing military UVs stateside, the same concerns arise regarding personnel and equipment on the ground.

A legacy of U.S. laws exist to dictate the probability thresholds for flight vehicles or debris unintentionally injuring people on the ground. Most of it stems from the world-war periods as the development of rockets and ICBMs intensified, followed by the space race. The Code of Federal Regulations Title 14—sometimes referred to as the Federal Aviation Regulations—requires that the probability of casualty for any single civilian on the ground not exceed 0.000001 per mission ($P_C \leq 1 \times 10^{-6}$). Expected total risk to the population must not exceed 0.00003 casualties per mission ($E_C \leq 30 \times 10^{-6}$) [203].

The military has adopted these requirements for test and training ranges to cover all airborne assets (airplanes as well as rockets, missiles, and artillery rounds). A thorough discussion of the regulatory history and justification for assuming the above limits can be found in guidance from the Range Commander's Council (RCC) [27]. This joint organization gathers and standardizes best practices for the operation of range and test facilities around the country. Quantitative requirements such as $P_C$ and $E_C$ limits are often met by test ranges using PRA based on vehicle component reliability data (if it exists), trajectory performance estimations, and population density calculations. This approach, developed in the middle of the last century, proves difficult when—instead of simple rockets or ballistic objects—the autonomous vehicles have wings and are capable of loitering, navigating, and decision-making.

Another undesirable event is two or more airborne vehicles colliding with each other. More complex factors than mere mechanical failure tend to be behind these types of incidents, and airspace deconfliction is a major focus area for the FAA and military [204], [205]. The DOD defines UAS categories in their Joint UAS Concept of Operations [206]. That document is confidential, but the definitions are not. Table 2-5, reprinted from the Army UAS Roadmap, summarizes the categories [18].

Table 2-5. UAS Group Definitions [18]

| UAS Category | Max Gross Takeoff Weight | Normal Operating Altitude (Ft) | Airspeed |
|---|---|---|---|
| **Group 1** | < 20 pounds | < 1200 above ground level (AGL) | <100 Knots |
| **Group 2** | 21-55 pounds | < 3500 AGL | <250 Knots |
| **Group 3** | < 1320 pounds | <18,000 mean sea level (MSL) | |
| **Group 4** | > 1320 pounds | | Any Airspeed |
| **Group 5** | | > 18,000 MSL | |

Different classes of flight rules and procedures exists for different altitudes and regimes of airspace. Typically, military UVs flying below 3,500 feet above ground level (AGL) are considered Groups 1 and 2 and are primarily operated by Army and other ground personnel who use the small vehicles—often deployed from their own rucksacks—for tactical support. Deconfliction with each other and MVs is purely visual, and collisions and near misses can occur [18]. At higher altitudes, UVs are equipped to conform to instrument and navigational capability requirements that allow them to integrate with the instrument flight rules that MVs follow. These Group 3, 4, and 5 UVs are typically operated from a remote GCS via line of sight or satellite link. They take off and land from conventional runways, and a lot of the focus for UV/MV deconfliction centers on the dense airfield traffic patterns. However, as UV density increases, the rest of the airspace will face challenges as well.

The AF operates three large-scale UV types: the MQ-1 Predator (Group 4), the MQ-9 Reaper (Group 5), and the RQ-4 Global Hawk (Group 5). The Predator, developed in the mid-1990s, was considered a highly useful asset once it was shown to be capable of employing air-to-ground weapons in combat, and it was put into mass production for use by the AF despite effectively being in an immature engineering development stage. Echoing historical trends in drone development, mechanical issues abounded, and eventually the Reaper was introduced as an enhanced evolution of the Predator. The Reaper is capable of roughly ten times the weapons payload of a Predator, uses refined manufacturing methods and materials, and improves on the human-machine interface in the GCS. Further improvements to the interface are in development to alleviate deficiencies noted by aircrew [207], [208].

The Global Hawk, developed in the late-1990s, is the largest of the AF UVs and began flying around the same time as the Predator. Its mission does not include weapon employment, but it has a large payload capacity in order to fulfill its high-altitude, long-endurance intelligence, surveillance, and reconnaissance mission. Whereas the Predator and Reaper UVs are manually remotely piloted in near-real time via data link, the Global Hawk is instead given navigational waypoints and objectives over its link. The Predator/Reaper GCS employs a stick and throttle, while the Global Hawk uses a mouse

and keyboard.[41] Degradation of the data link usually has more urgent impacts on the Predator/Reaper systems for this reason.

The following information through fiscal year (FY) 2014, was obtained from the *Washington Post* [19] and a search of public records[42]:

- In the DOD, there have been 194 Class A mishaps and 224 Class B mishaps (418 total) over roughly four million UV flight hours (average of 10.45 incidents per 100,000 flight hours[43]).
  - About one half of the mishaps happened in a major theater of operations (Afghanistan and Iraq), one quarter in a minor theater overseas, and the remaining one quarter in the continental United States.
- Almost half of the AF MQ-1 Predator inventory have been involved in a Class A or B mishap.
  - The five-year Class A mishap rate for all AF UVs during FY 2009–13 was approximately 4.3 mishaps per 100,000 hours, more than double the rate of the MV baseline (F-15 and F-16 fighters at 1.8 mishaps per 100,000 hours) during the same period. The five-year rate for UVs during FY 2010–14 improved slightly to 3.5 (the FY 2010–14 numbers for MVs were not available as of the writing of this thesis).
- The majority of major mishaps have been with the Army. It has lost 55% of its MQ-5 Hunter (Group 4 UAS) inventory and 38% of its RQ-7 Shadow (Group 3 UAS) inventory.
  - Army UVs had an accident rate ten times greater than its MVs in FY13 alone.

Appendix A summarizes historical accident statistics and accident investigation board (AIB) findings for the three AF UV types. The AIB findings are certainly eye-opening, but Sydney Dekker warns that publicly released accident data are not always an accurate source of information for the deep mechanisms of accidents [4]. The Air Force has a separate safety investigation board (SIB) process that has access to information that is legally privileged to only the flying community [35]. Often, those reports contain findings made in the light of practitioner culture, influences, and procedures, maintaining more user semantics. Regardless, with either type of investigation, it is very difficult to ascertain the hazards within a system by waiting for reportable accidents to occur, and even then, underlying factors might not come to light within the context of the system's design. Cullen thoroughly discussed how Predator and Reaper pilots work around deficiencies baked into the GCS design in his ethnography of that system [207].

---

[41] This is a gross oversimplification, as multiple computers, keyboards, and input/output peripherals will be encountered in any modern GCS. However, the Predator and Reaper cannot fly completely autonomously. A human pilot in the GCS exerts manual control via his flight controls (stick and throttle).

[42] See Appendix A for definitions of mishap classes and sources of public safety data.

[43] This is the standard format for reporting mishap rates in the aviation safety community.

Looking at the AIB information in Appendix A does reveal some trends worth noting, particularly in the mishap factor contributions. Most publically-available Predator and Global Hawk accident reports have cited some form of hardware or mechanical issue, while the vast majority (80 percent) of published Reaper accidents have involved HF (both during and pre-mission). This is perhaps not surprising considering the Predator and Global Hawk were first-of-their-kind UVs, and the trend has been to push the boundaries of software and control development while letting hardware maturity evolve naturally. The latest version of the Global Hawk is still experiencing airframe reliability issues [209]. These data are just from the releasable accident reports however. There are many scenarios for hazards that do not make it to the summary page of AIB reports.[44] They are instead tucked in the discussions mid-way through the reports, sometimes elaborated only in safety-privileged reports, or perhaps elucidated through in-depth investigations or ethnographies. The summary from top AF leadership, which also incorporates data from smaller scale incidents with no formal reporting, is that UVs have an accident rate six times that of MVs, and that 80 percent of incidents have HF-related causal factors.[45]

A systems view of safety will be useful when tackling the challenge of safe deconfliction of airspace for UVs and MVs. There have not yet been many notable airborne collisions involving UVs in the U.S., but a simple internet search yields multiple news stories of near misses in the civil and military domains. Military users themselves have developed issues trusting UVs [210]. The common factors the *Washington Post* cited found in historical UV accidents (most of them collisions with the ground) were human error, mechanical defects, unreliable communication links, and a limited ability of the pilot to detect collisions or undesirable positions/attitudes. One finding—not available through publicly disclosed accident reports but obtained by the *Post* through a freedom-of-information request—involved a UV being inadvertently flown upside-down [19].

As technology continues to push the boundaries of software capability, many aspects—data-link delays, difficult-to-read display symbology, or poor pilot awareness of UV orientation during approaches and landings, for example—must be interpreted as symptoms, not causes of accidents. While it is easy to push blame onto human operators when a "smoking-gun" mechanical failure does not exist (sometimes even when it does[46]), the lines of blame will become increasingly difficult to draw as more system responsibility is given to software. Software problems almost always stem from systems-engineering requirements flaws [5].

#### *2.2.1.2 Air Force Research Interests in Autonomy*

The AF has put much emphasis on developing future strategies for the development and implementation of autonomy-capable systems to perform and contribute to its core missions. The last generation of UVs were considered tools that could perform a limited

[44] Contributing-factor data in Appendix A were aggregated only from report summary pages.
[45] Dr. Mica Endsley, Chief Scientist of the Air Force, stated this during her address at the 18th International Symposium on Aviation Psychology in Dayton on 6 May 2015.
[46] When a human operator could have salvaged a mechanically defective aircraft through a procedure or best practice, accident reports often emphasize the fact.

set of well-defined tasks without human interaction. The future vision is for autonomy-capable systems to provide intelligent capabilities to the mission and operate in scenarios unanticipated by designers. Imagining how this might someday be possible includes understanding the capabilities that humans contribute to systems. Incorporating this knowledge with the system-theoretic view of humans in systems provides a baseline for the refinements that were made to STPA and presented in the next chapter.

*Technology Horizons*, a vision published in 2010 by the office of the Chief Scientist of the AF, describes the key technologies in which the AF must invest to meet its strategic capabilities over the next two decades. The vision considers autonomy a "disproportionately valuable" area [211, p. 3], and a proper implementation of autonomy is envisioned to enable practitioners to be able to focus on higher decision levels within their domain. People will be able to speak of capabilities instead of vehicles, of ad hoc instead of preplanned, and of adaptivity and resilience instead of defensiveness. The publication discusses key attributes of autonomy, including complex decision-making, unsupervised mission planning, and adaptivity to changes in the mission environment. Current AF UVs are considered to exhibit "limited autonomy" [211, p. 53].

The Defense Science Board convened a task force in 2012 to discuss the role of autonomy in DOD systems. To further autonomy development, the board recommended reducing focus from individual subsystems and warned against trying to label components as having particular levels of autonomy (see Section 2.1.2.1). Instead, they recommended the focus shift to software capability independent of physical platforms. They recommended allocating cognitive and decision-making functions among agents in a system, human or computer, and to understand the *collaborations* among agents. It also emphasized creating methodologies for tracing system behavior to system goals, so that during design and operations the tradeoffs among top-level properties (e.g., performance, efficiency, manpower savings) are evident [212].

The DOD, as well as the Army and Air Force, published UV road maps for the next several decades. Although these documents refer to levels of autonomy—they were all written before the Defense Science Board findings—the top-level thrusts are common. They all advocate for the increase in flexibility and adaptivity that improvements in autonomy can provide. Other common visions include dynamic environment capability, persistence, interconnectivity, and usability across domains [18], [213], [214].

AFRL has been tasked with advancing software capabilities as well as verification and validation (V&V) methods for complex systems. Their 2013 autonomy strategy contains a simple vision: "Intelligent machines seamlessly integrated with humans maximizing mission performance in complex and contested environments" [15, p. 4]. Objectives include advancements in human-machine teaming, shared perception, self-governing teams of machines, robust communication, and flexible decision-making paradigms. A large emphasis is put on streamlining software certification requirements when portions of systems change or recombine for different tasks. "Preventing unintended emergent behavior...and maintaining safety guarantees at the system level" become important, and modernized test and evaluation strategies are called upon to make high-level determinations [15, p. 13].

AFRL also emphasizes that "automation" and "autonomy" have different meanings, admitting that in the past the terms were often used interchangeably. They offer definitions that are based on a machine's capacity for decision-making abilities [15, p. 3]:

*Automation* – System functions with no/little human operator involvement; however, the system performance is limited to the specific actions it has been designed to do. Typically these are well-defined tasks that have predetermined responses (i.e., simple rule-based responses).

*Autonomy* – System has a set of intelligence-based capabilities that allow it to respond to situations that were not pre-programmed or anticipated in the design (i.e., [knowledge][47]-based responses). Autonomous systems have a degree of self-government and self-directed behavior (with the human's proxy for decisions).

While the purpose of this thesis is not to investigate the semantics of the various definitions of automation and autonomy[48] that exist, some emphases merit discussion. *Automation*, taken in the simplest sense to mean function without (or with little) human involvement, is evident in a typewriter from the 1870s [14]. Although it takes more complex forms in things like high-speed elevators, car-manufacturing factories, and airplane cockpits, the concept is the same. Kathy Abbott at the FAA highlights that data sensing, information interpreting, deciding, actuation, and/or any combination thereof can be automated with computers, equipment, or machinery [215]. This is important because it parallels the STAMP view of control loops and the sections of those loops. Something like a healthcare information-management system would be an example of the automating the information-interpreting function. The typewriter would be an example of automating the actuator function, where the loop being examined is document-author-document. The original Fitts MABA-MABA list, presented in Section 2.1.2.1, compared humans and machines only in the narrow context of each of those control loop sections.

By moving from an emphasis on automation to an emphasis on autonomy, the DOD is showing a shift from technology-centered thinking to a work-centered approach. In order for *autonomy* to be possible as defined by AFRL, all functions (sense, interpret, decide, act) must first be automated in all the control loops within the intended autonomous system. While complete automation is necessary it is not sufficient, however. AFRL states the ability for a machine to sense, perceive, plan, decide, and act requires *machine intelligence*, citing Visnevski and Castillo-Effen [216]. How an intelligent controller in a system is able to interpret domain information and make decisions is important. The following excerpt from the autonomy strategy is notable [15, p. 14]:

> For a machine to perceive its environment, it must not just sense it but also be able to extract information. Planning involves task development, sequencing, and future

[47] The original word here was "decision." It has been replaced with "knowledge" to align with terminology coined by Rasmussen in his discussions of skills, rules, and knowledge-based decisions (S-R-K framework) [165].

[48] "The attempt to define autonomy has resulted in a waste of both time and money spent debating and reconciling different terms and may be contributing to fears of unbounded autonomy" (Defense Science Board) [212, p. 23].

> state prediction, which will require significant advancements in machine intelligence. Teaching a machine how to select an option, act, and then validate that its selection and action is appropriate is a key part of demonstrating autonomous technologies. For machines to become intelligent, they must have the ability to learn and adapt state, knowledge, behaviors, decision-making processes, and teammate interactions through learning. Knowledge representation and transfer (symbolic and sub-symbolic reasoning) are key areas to develop and mature machine intelligence. The ability to detect, isolate, and reconfigure due to faults means a system can better perform on its own and in concert with other team members.

Industries investing in autonomy including the DOD are pushing towards hierarchical-control concepts that will go beyond centralized-control-decentralized-execution to include decentralized decision-making at lower levels, allowing quicker system adaptations [16]. Flatter decision responsibilities do not reduce the need to communicate intent and constraints down through the hierarchy, but increases it [217]. Non-reductionist approaches to understanding system designs must consider not just the behavior of components, but the possible relationships between components as well as their local interpretation of system values. The mapping of shared goals and motives to lower-level behavior must be understood by decision-makers during operations for a system in order to handle dynamic goals in dynamic environments without the risk of being brittle or non-adaptive.

While there are numerous meanings for "adaptive" depending on the domain involved, the following terms are considered [218]:

*Optimized* – System can satisfy fixed objectives in a fixed environment.

*Robust* – System can satisfy fixed objectives and adapt to changes or uncertainties in the environment or the system itself.

*Flexible* – System can also adapt to changes or uncertainties in objectives.

It follows that an intelligent controller (be it human or machine) requires the capacity to sense and interpret domain information specific to the level of adaptivity they contribute to the system. A controller designed for optimization may only have to monitor a few specific process variables and be well-designed (or trained) to match a reference signal within acceptable tolerances of accuracy and dynamics. A controller that contributes to robustness might have several control strategies and the capability to utilize or cooperate with other controllers in the hierarchy. A controller able to impact system flexibility is capable of understanding, communicating, and prioritizing higher-level values of the system. Robustness and flexibility, the latter in particular, are marked by controllers that are not only optimizing specific tasks but more primarily satisficing perceived system goals [45], [68].

TRLs are defined to develop new scientific discoveries that are demonstrable with experimental rigor and develop them into functioning technological components that can be tested in isolation. Meeting the requirements of TRL levels (3) through (7) requires constraining a finite opportunity-space and sanitizing the testing to exercise specific corner points of the performance. As systems become more complex and software-intensive, the

number of potential combinations of states will quickly outnumber designers' and testers' ability to predict regions of unknown or undesired behavior, making specifications undefinable and risk reduction non-exhaustive of the problem space [15].

Enabling more capable autonomy will require more powerful methods of software V&V [38], [53], [91]. The AF has acknowledged that ensuring expected behavior by software is not possible by trying to simulate the near-infinite number exclusive states. New methods of V&V are therefore called upon to certify future systems, in order to keep up with competitors who might not put the same level of scrutiny on their developments [211].

According to the *Defense Acquisition Guidebook*, verification ensures that a "system or system element performs its intended functions and meets all performance requirements," while validation ensures the "capability provided by the system complies with stakeholder requirements, achieving its use in its intended operational environment" [23, pp. 329–330]. In other words, verification asks *does it do things right?* while validation asks *does it do the right thing?* In typical systems-engineering developments, V&V cycles are embedded within many stages and sub-stages of a program, with each one satisfying the performance specifications are met and then confirming that higher level requirements are satisfied [46]. Once the possibilities for behavior states outnumber the capacity to evaluate all of them, the bottom-up mapping from performance to system goals becomes very difficult to validate.

Safety, as a system goal, is just as difficult to verify if done from the bottom up. STPA enforces safety from the top down by enforcing constraints through a hierarchical control structure. The refinements to STPA discussed in the next chapter extend the method to capture more scenarios involving humans and autonomy. When STPA is then applied to product testing, it has more appropriate tools for investigating these modern systems.

### 2.2.2 Developmental Test

Industries with comprehensive systems-engineering processes do not simply put new systems in the field to evaluate them merely based on initial use.[49] Responsible testing of products often begins in the research and design stage, with formal testing in the form of lab or bench tests. Components or assemblies undergo rigorous and repeatable functional evaluations in a controlled setting. In the AF, developmental product testing becomes prevalent during the transition from TRL (6) to (7) and beyond. At this stage, programs exist on record, and the systems-engineering process has developed extensive system and subsystem requirements and specifications that must be verified. AFTC is an entire enterprise devoted to test within the AFMC organization. This cadre of professionals is responsible for building sanitized reproductions of the field-use environment in order to evaluate transitional systems in close-to-real-world conditions.

Chapter 4 goes over AF DT in more detail. The goal of DT is to "demonstrate systems feasibility, confirm engineering design and development are complete, minimize design

[49] A mechanic who tinkers with his own car engine with the expectation of observing greater output when driving on public roads would be an example of testing only in the field.

risks, and ensure systems perform as required in their intended environments" [219, p. 1]. The goal can be stated as two parts, which are not necessarily mutually exclusive:

- Specification compliance within a representative environment
- Risk reduction (for technology capabilities, safety, performance, etc.)

An example of specification compliance would be making sure a radar beam can discriminate a certain-sized target at threshold and objective distances. Risk reduction might be found in sorties performed to explore a useful envelope for a new airplane modification or capability upgrade. Verification is accomplished in DT by using a sanitized testing and measurement framework to isolate the item or capability being evaluated. The intent is internal validity and repeatability. The DT stage also involves extensive modeling and simulation in order to predict results before test conduct. Once testing commences, results are corroborated with predictions. Maneuvers are designed to start off benign and gradually increase (during the same test or in subsequent operations). This is known as the *buildup approach*; it breaks up risk and uncertainty into smaller pieces [220].

Testing is ultimately about reducing uncertainty, whether it is performed to meet a system specification, lower design risk, or both. In any case, certain assumptions are made before the testing based on designs, models, lab experiments, or simulations. DT personnel sometimes refer to AFTC as the "tip of the whip."[50] Despite efforts to involve AFMC in the using-commands' requirements-development processes, the DT stage might not always understand the entire reasoning for the design and specification decisions that have been made before a test product arrives at a DT squadron. Without that framing and with extraordinary schedule pressure being the norm, the test profession is still capable of designing tests and experiments in order to maximize efficiency while accomplishing enough specification compliance and risk reduction to justify the initial fielding of a product.

MIL-STD-882 states the objective of safety during test is to "eliminate [or reduce] the hazards for both the system and the test [unique] events" [11, p. 82]. The following explains those two goals:

A) Determine safety of the system as designed and intended for use[51]
   - Confirmation that design-stage models (e.g., computational dynamics) used adequate assumptions and input parameters
   - Risk reduction for aspects of the system with no accepted models (e.g., human or autonomous components)[52]

B) Ensure safety of the testing process itself

[50] Based on fact-finding interviews I conducted at Edwards AFB.
[51] Along with verifying the compliance with certain specifications, the AF uses an airworthiness certification process by which to approve a system as safe to fly (more in Chapter 4).
[52] Even when uncertainty of human or software behavior is acknowledged, testers will still apply assumptions about field procedures when exercising the system. Those assumptions need to be documented just as any regular design assumptions.

- Test techniques, configurations, instrumentation, range support
- Buildup approach when verifying models that might be inaccurate

Product testing must both evaluate acceptable system safety (as defined by the program managers) and account for risk to testers, range/support personnel, and the public. The second goal, *test safety*, is a unique consideration during DT [24]. The general programs of *flight* and *weapons* safety—which complement TTPs in training and fielding—are not enough during testing. Two driving factors for test safety exist: a) there are product configurations, instrumentation, techniques, and maneuvers that are unique to testing; and b) the inherent safety of the design itself is not yet proven.

Initial fielding follows DT and almost always begins with end-user evaluations. The goal of field evaluations is to "demonstrate, under as operationally realistic conditions as possible and practical, that systems are operationally effective, suitable, and capable of meeting the user's requirements" [219, p. 1]. The intent is ecological validity and system behavior under real world uncertainty. This testing is done by field users using established TTPs as much as possible. These field testers are either operators assigned to the using commands, or they come from an independent organization called the Air Force Operational Test and Evaluation Center. This direct-reporting unit is a congressionally-mandated independent-evaluation organization that reports to AF headquarters. It evaluates programs with high cost and impact thresholds, and its operators must perform an initial field evaluation before full-rate production is authorized on the product [194].

During field evaluation and use, operators still follow traditional flight and weapons safety programs. Although test safety as an independent planning consideration does not exist in the field, initial use-limitations might exist due to the results found during DT. If field users encounter a problem with the design or behavior of the system or one of its components, they can communicate this through a problem report. A type of problem report called a *deficiency report* (DR) is part of a mandatory program for government acquisitions; DRs provide manufacturing, testing, and field-use practitioners a means to document quality assurance (QA) problems [13], [221].

Figure 2-12 summarizes the product stages and highlights important segments within each stage, as well as tasks that are performed on a product. While DT aims to reduce uncertainty in the system design, field evaluations aim to reduce uncertainty in the system's emergent behavior when used in a real-world environment. Another way to view the difference between DT and field evaluation is in the light of verification and validation, as defined in the previous section. Although V&V loops exist within all stages of system development, on a grand scale DT can be viewed as a verification stage for individual system components and their behavior, while field evaluation serves as the initial stage for validation that the system is usable and meets design intent. Efforts to combine DT and field evaluation have been prevalent in the last decade as development budgets have decreased and product complexity has increased. It is not rare to find using-command members assigned to liaison positions in AFTC squadrons, often operating test products alongside DT operators while accomplishing field evaluation items.

With continued research and development of autonomy, the ability for the DT enterprise to make determinations about how behavior traces to system goals might be as

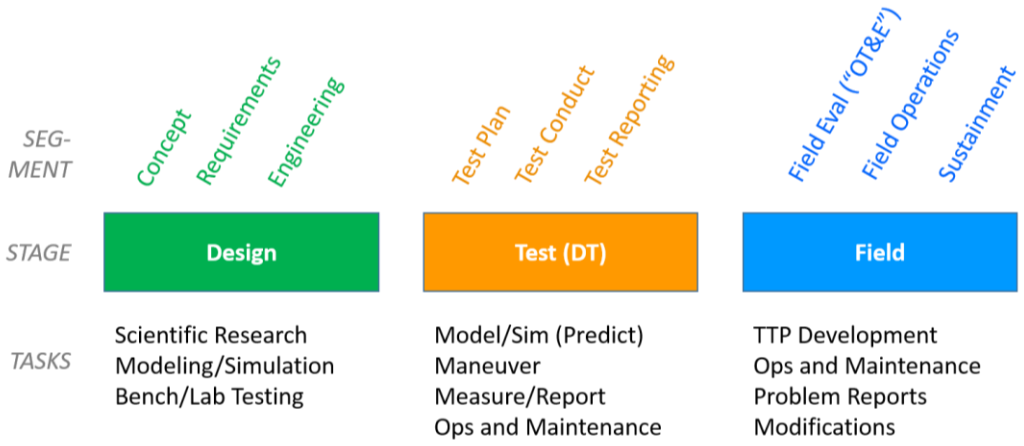



Figure 2-12. Product Stages

important as component verification. Although the DT stage is not expected to validate high-level performance goals of the system, high-level safety properties of the system must be a concern to anyone operating it. Methods for understanding how safety emerges in the system will naturally yield insight into how mission performance emerges. Having that ability in DT will require a more consistent communication of a product's functional framework, including its design and use concept, from the using commands to the program managers and then to testers. DT testing of autonomous systems might lend itself to ecologically valid conclusions earlier in the development cycle because it directly tackles the performance and safety issues that will be found in the field.

Missing from DT—but more importantly, the entire acquisition process—is the ability for all stages to share a common model for safety. A common framework would have multiple benefits. It would provide an organized method to document and trace assumptions and acknowledged uncertainties about the design and/or controller behavior. This would allow testers to be explicit about the focus of their activities and reporting. It would also allow findings during DT to be put into the context of envisioned field use. It would give testers a systems-based method to *visually* plan for the safety of the test process. Finally, it would treat safety as a single emergent property of the entire system, allowing individual safety disciplines (e.g., flight, weapons, test) to be considered under one approach.

Man-made systems are indeed growing more complex, requiring directed efforts to design, build, test, field, and manage them to achieve their purpose. Safety exists at the top level of a system, and like other system-level properties, must be understood and implemented with control-based methods using the viewpoint from systems theory. Neither software behavior nor human behavior can be reduced to mere physical phenomena when they are coupled within dynamic work ecologies. STPA can be refined to capture more human behavior, as presented in the next chapter. Testing systems involves not only ensuring that they are safe to field, but building a strategy to ensure that the test activities are safe. STPA is incorporated into test planning in Chapter 4.

# Chapter 3

# STPA Considerations

> "People approach the world not as naive, blank-slate receptacles who take in stimuli as they exist in some independent and objective way, but rather as experienced and sophisticated veterans of perception..."[1]
>
> —*Deborah Tannen*

The first research objective is to *extend System-Theoretic Process Analysis (STPA) to better examine human controllers in the hazard analysis*. The following are the existing gaps in the STPA controller analysis prior to this research:

- The *process model*, one of the main concepts of STAMP, does not capture types of system abstraction that human controllers require to contribute adaptivity to system behavior.
- Fundamental human considerations are not explicitly considered in the controller analysis (e.g., workspace factors, variability of personal traits).
- There is no current method to model the impact of social and organizational influences on the controllers within the operating process.

I performed this work at MIT. In order to understand the existing guidance for analyzing human controllers in STPA, I reviewed past STAMP models and analyses for format and content. I developed an extended analysis to add granularity to the controller mental-model analysis, detail fundamental human-engineering considerations, and consider the influences to controllers that evolve prior to a system operation. I developed a visual tool to document some of those influences. In order to examine the utility of the STPA extension, I compared results of an analysis performed on a real world concept using the extension with previous results using STPA without the extension. Results are tabulated and contrasted along the logical partitions of the updated method.

## 3.1 Intelligent Control

Humans play an important role in accidents (both positive and negative)—as demonstrated by the unmanned vehicle (UV) incident findings discussed in Appendix A—and they must

[1] *Framing in Discourse* [222, p. 20]

be included in hazard analyses. As discussed in the previous chapter, theory and techniques were developed in the human-performance communities to explain and measure mental and physical limitations, and the functional allocation of work tasks based on the typical performance capabilities of humans and machines has been a common design approach. However, the various standards, specifications, and checklists for human-engineering design cannot prescribe solutions that account for the individual qualities of any specific design or context. Applying universal guidelines and best practices without a consideration for design tradeoffs between performance, usability, and other requirements might cause unintentional conflicts with safety.

While functional allocation has its place—and simple or common human errors can be avoided with well-informed design principles—there is a larger source of human contribution to accidents that is all too often treated by designers only with guidelines, standards, and approaches that are reductionist. Current hazard analyses and safety investigations, in similar vein, discuss human contributions to accidents in a simplistic way, if they include them at all (e.g., pilot failed or pilot lost situation awareness[2]), and then they assign a probability to this failure. Leveson states [5, p. 273]:

> Humans do not always follow procedures, nor should they. We use humans to control systems because of their flexibility and adaptability to changing conditions and to the incorrect assumptions made by the designers. Human error is an inevitable and unavoidable consequence.

Humans in systems are a source of safety, not just accidents. Today, the differences between a person and a computer go beyond the superficial traits of physical and processing power. The degree and predictability in which a human considers different abstractions of the operation/mission, explores for new data, attempts unique actions, and taps into various modalities of information are important to consider. This requires a system-theoretic approach to finding discrepancies that could lead to accidents. Focus should be emphasized on the responsibilities of (and influences on) humans within the design as well as the use philosophy of the system. This goes beyond merely stating that a controller failed.

Humans can dynamically adapt their control strategy depending on the systems they are within [72], [190]. This adaptivity was first evident in findings from manual-control research, in which it was found that when a person was put in mechanical series with different actuators of varying transfer functions, the person would adapt their motor response to create a combined human-machine closed loop system with a transfer function that emulated a simple integrator [75]. The dynamics of the human control algorithm are also evident in more complex types of controlling, supervising, and decision-making. Humans creatively seek feedback from multiple sources in order to decrease uncertainty in

[2] Dekker points out that situation awareness is not something a person loses; they always have an interpretation of the system states [4].

their mental models, and change the structure of the mental models themselves to accommodate new types of information [21], [223].[3]

In seeking to achieve goals, human controllers go beyond trying to maximize behavior only for local objectives. They seek to satisfy many system motives (including safety), which requires constantly valuating their importance, estimating what other controllers' priorities and control strategies are (when that information is lacking), and exploring unique optimizing techniques within their control influence. What people take for granted as sound judgment is an intricate assortment of these motive-balancing decisions. Government range-safety officers, for example, sometimes have a unique responsibility to terminate flight vehicles (manned and unmanned) via destructive devices should the vehicle violate certain airspace boundaries [224]. When the vehicle in question is new or unproven, it is often left to the discretion of this officer whether a slight incursion or near incursion into the airspace boundary is the only parameter that should warrant destroying the one-of-a-kind and/or expensive system while the vehicle is making perceptible efforts to correct itself.

STPA in its current form is a powerful hazard analysis technique because it considers behavior of the system in the context of its functional design, and it starts the safety analysis with well-defined system constraints. Because of the increase of interacting humans and autonomous components in modern complex systems, refinements to how an STPA analysis handles human controllers also helps identify more hazardous scenarios. For the human, the STPA controller analysis currently looks for discrepancies in the feedback, mental model, and decision-making. This method, although advanced compared to the traditional safety analysis techniques, still oversimplifies the human's role in complex systems because it is not different than examining a machine controller's process model and control algorithm. Human mental models contain more information about the system than a machine's process model, and mental models develop using more sources of feedback. In addition, the performance constraints and variability of humans are valid engineering considerations that should be incorporated into the analysis. Finally, influences on the controller's decision-making activities evolve prior to the operating process (see Figure 2-7 in the previous chapter). Like human-performance considerations, these aspects should be captured by STPA when generating causal scenarios.

By addressing the gaps in the STPA controller analysis, I developed an extended analysis, generalizable to any controller, not just humans. Advanced process models will eventually be a characteristic of autonomous controllers as technology continues to push the capabilities for software to be adaptive. This would make the autonomy an *intelligent controller* just as much as a human. An intelligent controller has the capacity to tradeoff and satisfy various system goals and parameters beyond the foresight of designers [15].

---

[3] An example of seeking unofficial sources of feedback would be an office worker calling a friend in another department to get a business forecast so he can begin to prepare his workflow to accommodate it. Accommodation, also discussed in the previous chapter, might be demonstrated by an experienced city driver who takes the tunnel on their commute; she has learned that on a busy-traffic day, she can look for a slight reflection of other vehicles' brake lights on a small portion of the tunnel wall which warns her to slow down before rounding a curve and seeing the actual cars. A brand new driver would only rely on the sight of the real cars once he makes it around the bend.

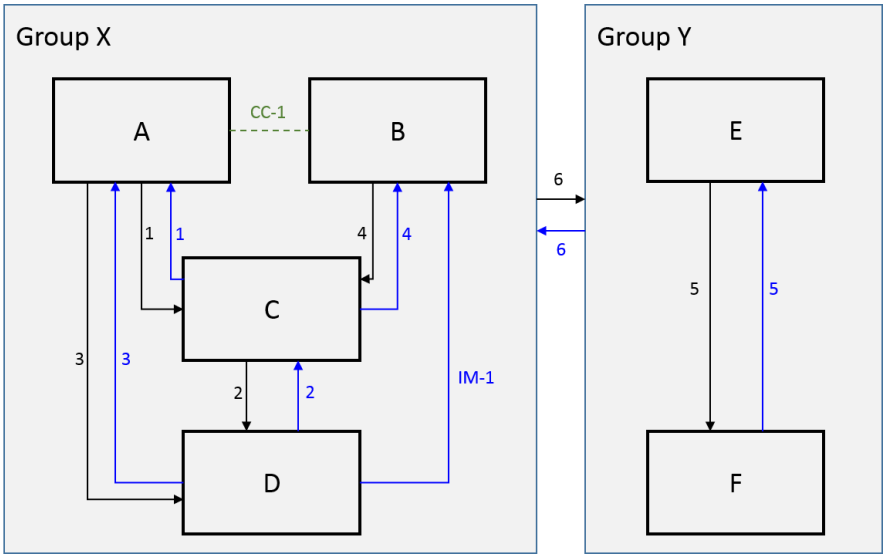



Figure 3-1. Control Structure Format

Additionally, influences from outside the operating process can affect any controller. The new extension ensures that all controllers, current and future, may be examined in the hazard analysis.

The next introduces the format chosen for this research for constructing the System-Theoretic Accident Model and Processes (STAMP) control structure for use throughout the thesis. Section 3.3 presents the new STPA extension in the form of new visual aids and additional guidance for the controller analysis, followed by a brief real-world example in Section 3.4 to demonstrate findings using the extension to the original STPA technique. The next chapter assumes these conventions and introduces additional concepts related to systems testing.

## 3.2 Visual Format

Figure 3-1 demonstrates the visual format used in this thesis for the safety-control structure. Introduced in the previous chapter, the safety-control structure illustrates the relationships between functional levels of control in the system. It is not necessarily a representation of the system's physical structure, nor is it simply a design schematic or an organization chart. However, creating it does require knowledge of both the physical design and the information relationships among controllers (human and machine). Control relationships are generally illustrated in this model such that higher controllers have broader responsibility, authority, and/or accountability.

The entities shown in a safety-control structure are part of the system of interest, and by definition they are located *within* the system boundary and can be influenced by design and operations. The *environment* is not illustrated; it is considered explicitly later on in the analysis. The example in Figure 3-1 contains six process entities (A–F). Each one

can be a controller, a controlled process, or both.[4] The entities would normally be labeled what they really are (e.g., pilot or flight computer) instead of letters. The primary relationships between higher and lower entities are control actions (CA) and feedback (FB), shown as black arrows and blue arrows, respectively. Two-way communications between entities (CC) are shown as green dashed lines. Another type of relationship is the indirect measure (IM), shown as a blue arrow because it is typically from a lower to a higher entity but does not meet the criteria of feedback.[5] Each communication relationship—CA, FB, CC, and IM—in the figure has been labeled and assigned numbers for reference. CAs and FBs are numbered in the figure but do not include their respective "CA" or FB" labels in order to save space and for readability purposes.

There is no fixed shape that must be used for all control structures. Although Figure 2-7 in the previous chapter might suggest that they should mostly look like ladders,[6] this is not always the case as Figure 3-1 shows. Control and feedback paths can skip across levels in the control structure, horizontal relationships may exist, and multiple controllers can operate at the same level. The control structure must be intricate enough to model the system in order to focus on the entities that engineers wish to influence with design and operating recommendations. By making functional control the focus of STPA, analysts may remain abstract in some parts of the model. For example, an operational safety analysis of an existing aerospace design might model an aircraft as a single entity (black box), while a detailed design analysis performed earlier might break that entity up further into the functional components within the aircraft.

Sometimes it is useful to show part-whole decomposition within an entity but maintain the concept of that entity. In Figure 3-1 this is demonstrated by the two groups X and Y, which have a control-feedback relationship between them ("6" in the figure). This could mean that both Controllers A and B may issue commands to and receive feedback from Controller E, for example. An adequate description of the control relationships should accompany any system's model before any hazard analysis begins.

Imagine Figure 3-1 represents a manufacturing plant. The control structure describes an operations manager (A) overseeing a line supervisor (C) of the assembly line (D). The operations manager also exchanges information about product requirements with a sales manager (B). Each of the connections in the control structure (e.g., CA-2) might have one or more variables (e.g., 2a and 2b) that contain different information content. A

[4] The lower entities in this case (D and F) are not controllers, and are typically (but not necessarily) the primary interface between the system and the environment. It is also implied (but not shown) that all entities can each have a process input from the environment, a process output to the environment, and disturbances from the environment.

[5] Feedback to a controller comes only from entities it directly controls; if a controller is using information about another entity it does not directly control, that information is an indirect measure. Similarly, CC relationships can be considered as two-way measures, with similar information in both directions.

[6] A ladder in a control structure is a single vertical column of entities, each having a CA and FB relationship only with the entities below it and above it.

Table 3-1. Example Variable Reference

| Variable | Name |
|---|---|
| Control 1a | Desired Output Rate |
| Control 1b | Break Schedule |
| Control 2a | Conveyor Speed |
| Control 2b | Item Spacing |
| Control 3a | Inspection Rate |
| Feedback 1a | Output Rate |
| Feedback 2a | Conveyor Speed |
| Feedback 2b | Output Rate |
| Feedback 3a | Inspected Results |
| Feedback 3b | Output Rate |
| Comm 1a | Product Requirements |
| *etc.* | … |

simple variable reference can be built to describe each labeled connection between entities. An abbreviated example is shown in Table 3-1.

It is important to always consider informal as well as formal relationships between entities when forming and updating the control structure. This is one of the reasons why documented activities and problem reports are useful items to supplement official schematics and organization charts. More rigorous methods, like ethnographies, might be used to better understand the connections within a complex sociotechnical system.[7] The control model should be periodically addressed to identify new relationships. Section 3.3.1 presents a step to encourage control model updates during the analysis (non-designed feedback).

### 3.2.1 Phases and Subphases

Besides broader responsibility, authority, and accountability, another common characteristic of higher-level control loops in the STAMP model is a longer time constant [185]. Said another way, the lower in the control hierarchy an entity is, the higher the rate at which its local information and activities tend to function. In the manufacturing plant example, the line supervisor might monitor and adjust the assembly line parameters every fifteen minutes, while the operations manager might check on the line supervisor only every two hours.

Every development stage—or stage segment (see Figure 2-12 in the previous chapter)—of a product's life can be modeled with a safety-control structure. At the bottom

[7] Cullen's review of Reaper operations revealed a concert of intra-communications and data referencing techniques happening in the UV ground station, including various references to technical manuals, mission cards, computer databases, and internet chat [207].

the control structure can be one or several *phases*, which are *defined periods or units of work activity*. Phases are socio-technical activities that may repeat or alternate with other types of phases (e.g., plan, brief, operate, debrief, repair, etc.). A phase operates at a short time constant with respect to the higher levels in the control structure. The control-structure example in Figure 2-7 in the previous chapter shows a phase in the system-fielding stage called the operating process; however, other phases, such as maintenance, exist as well. The system-development stage would contain phases such as design and manufacturing.

Enterprises should identify and acknowledge each type of phase in their work and what the phase's typical time constants are. In fielded commercial aviation, an operating phase might be a single flight (e.g., two to eight hours), while a maintenance phase might be one to two weeks of repairs. For continuous work, like in a nuclear power plant, an operating phase could be defined as the eight-hour control-room shift of a reactor supervisor, or perhaps one to two months of power production between planned maintenance shutdowns.

The identification and acknowledgement of phases is important because activities that occur in one phase can have impacts on the behavior of a subsequent phase. A simple and intuitive example would be that of a maintenance technician making an inappropriate adjustment to a system component that manifests as a hazardous behavior once the operating phase begins. Another example is a worker that forgets to reset his display settings before handing operations off to the next worker, and the new worker might then read the displays incorrectly. Some of the "pre-mission considerations" discussed in Appendix A that contribute to Air Force (AF) mishaps are also examples of influences between phases.

Often, hazard analyses like STPA are performed on only one phase of a product (e.g., the operating process in Figure 2-7 in the previous chapter) and usually only in the fielding stage. While this thesis discusses the test stage in the next chapter, it still focuses mostly on the operating phase within that stage. A complete safety-control structure for a given stage would ideally contain the entities for all the possible types of phases. For example, if an aircraft were located at the bottom of the system-fielding control structure, both a pilot and a maintenance technician would be entities above it. However, control structures might be simplified to only focus on one type of phase. This is another reason to have documentation identifying and acknowledging the other phases in the enterprise. Section 3.3.7 introduces methods to look for influences that other phases have on the phase being examined in the hazard analysis.

*Subphases* are defined as *sets of related behaviors within a phase*.[8] An example of subphases in the operating phase of fielded commercial aviation would be startup, taxi,[9] takeoff, climb, cruise, descent, landing, taxi, and shutdown. Subphases are important to identify and acknowledge because controllers might change their gains, assumptions, and

[8] This term is defined similarly but called controlled-system operating mode (or state) by Leveson et al. [38].
[9] Taxiing is the act of using engine power to propel an airplane along the ground at a slow to moderate speed to maneuver it to different locations while on the ground.

control algorithms depending on what portion of the phase they believe the system to be in. More is discussed in Section 3.3.3.

Systems engineering is an iterative process, and the safety-control structure should be updated by engineers as discoveries and modifications are made on the system. Hazard analysis is a human activity, and by emphasizing the maintenance of consistent models like the control-structure and its variables, analysts will regularly question their assumptions about the content of these products. Once the control structure exists and accidents and hazards are identified, the ensuing STPA hazard analysis identifies hazardous behavior in the form of unsafe control actions (UCA) as discussed in the previous chapter. Tables are one method for organizing UCA findings, and multiple formats exist for constructing those tables, some based around formal methods for identifying UCAs developed by Thomas [189]. The four-column method already introduced earlier is used for this thesis. The remainder of this chapter focuses on extending STPA.

## 3.3 Proposed Extension: STPA-RC

I developed a method to extend STPA called STPA-RC, where "RC" means *refined controller-analysis*. The remainder of this chapter discusses its philosophy and provides guidance for its implementation with new visual aids and additional considerations for the controller analysis.

Incorporating human behavior into STAMP and STPA has taken several forms in previous research, and STPA-RC aims to capture the important attributes of those efforts and address the existing research gaps to produce an updated analysis technique for intelligent controllers. Before introducing the new material, a brief overview of previous work is appropriate.

Leveson began developing controller models with human considerations predating STAMP. Figure 3-2 reproduces an early version of the human controller in a control-theoretic format [5]. Leveson stated that the human controller is often managing automation, which is controlling a lower process. The human is thus indirectly controlling the process, while occasionally there is direct control and feedback between the human and the process. The human process model requires information about the automation, shown in the figure as a separate mental model. The figure also includes an additional process model entity to account for the human's understanding of the situational context of the system.[10] Extrinsic factors are mentioned but not elaborated further, and a control-action generation component exists that includes decision-making and action initiation.

Thornberry modeled human controllers as reproduced in Figure 3-3 (blue circles are my emphasis) [190]. He expounded upon the extrinsic factors to include considerations like written procedures, environment, and culture. He also argued that human-sensory perception—aspects of the physical environment that are directly sensed by humans, such as vestibular forces and optical motion cues—should be explicitly considered as a distinct

[10] The context could, for example, be the subphase of the operation the system is in or the particular environment in which it is functioning.

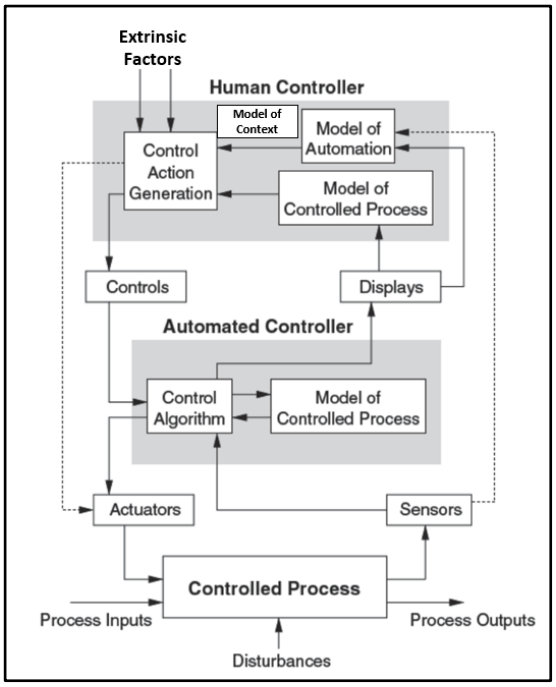



Figure 3-2. Human Controller Model, Original [5]

feedback because design engineers may not consider it as an official source of feedback when developing the safety-control structure for the system. He added a section for detection and interpretation, acknowledging that the presence of feedback is not a guarantee that the human controller perceives the feedback. Thornberry also divided control-action generation into two distinct sections of decision-making and affordance. This resulted in four intrinsic parts of the human-controller model (detection, process model, decision-making, and affordance) as well as a pooled set of extrinsic factors.

Affordances are a concept Thornberry borrowed from Gibson and later work by Flach, based on the system-theoretic viewpoint of humans in systems [153], [154]. Assuming a constructivist perspective of cognition, an affordance is a human controller's ability to perceive available opportunities and consequences and generate actions in the particular work ecology. It is where the effectiveness of the agent's mind couples with the possibilities of the work, causing humans to act according to what makes sense to them [183]. Affordance is implied in the "detection and interpretation" section of Thornberry's model, and it is explicitly denoted where control-action generation formerly resided [190].

Another influence to Thornberry's human-controller model update was the work of Boyd, a military strategist who developed a cognition model called Observe, Orient, Decide, Act (OODA) [225], [226]. Thornberry acknowledged Dekker's warning about the inability to prove or disprove models of cognition [4]. However, Boyd was a constructivist thinker, and the appeal of the OODA model is that it describes dynamic decision processes from a systems perspective [16]. The activity of observation (detection) influences and is influenced by the orienting (process model) activities, and vice versa. The activity of deciding influences observing and orienting. When actions are afforded onto the work

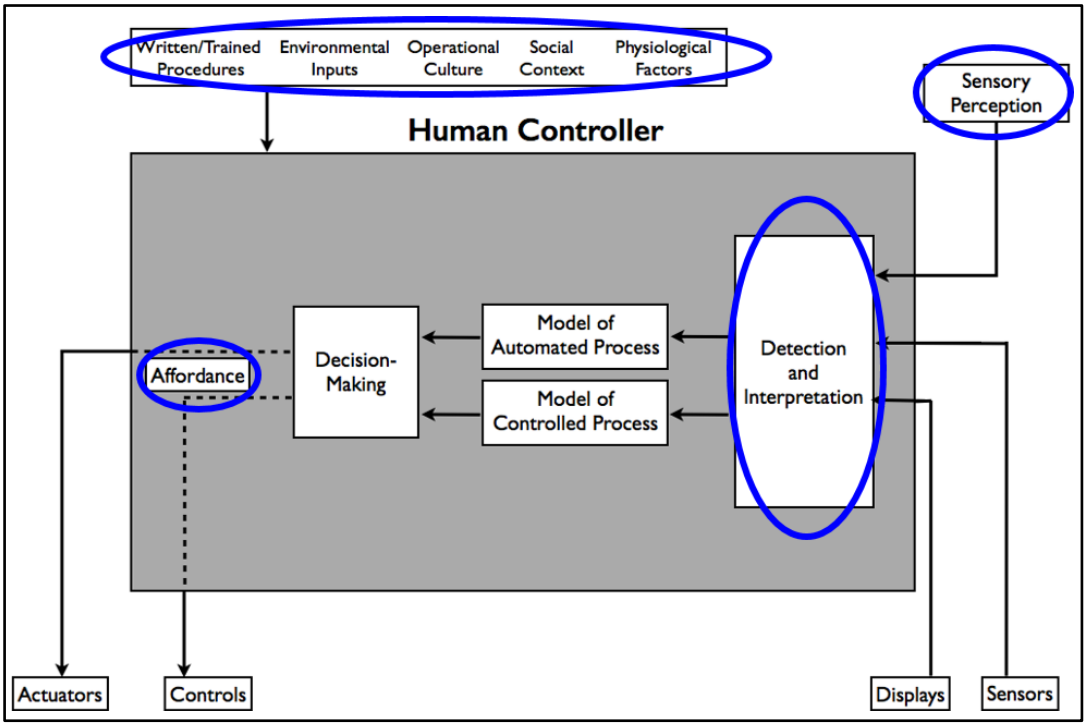



Figure 3-3. Human Controller Model, 2014 [190]

ecology, their effects are searched for and observed. The OODA loop is in essence a decision-making perspective for complex activity that acknowledges the adaptivity of human controllers. Although Boyd's original figure is not reproduced here, its four key areas are evident via the four intrinsic sections of Thornberry's model.

Stringfellow developed a human-error taxonomy for organizational human factors; it offered hazard-analysis guidance that mapped to the parts of the basic STAMP controller model (i.e., feedback, process model, and control algorithm) and should be considered during an analysis similar to the ones in Figure 2-9 in the previous chapter. The guidance included items like "goals wrongly prioritized [control algorithm]," and, "inadequate understanding of process boundaries [process model]" [21, p. 108]. Thornberry also developed analysis guidance, providing considerations for causal scenarios based on his model in Figure 3-3. STPA-RC is itself an updated analysis; it does not require updating any models of the human controller, but rather it acknowledges existing models and maps to their components.

A visual representation of STPA-RC is shown in Figure 3-4. It is an analysis technique and not a cognition (or information-processing) model for reasons already discussed. It does not mention *automation* per se because, from a systems perspective, STPA is based on the functional relationships and behaviors of all entities in the hierarchy, regardless of whether they are machine or human. STPA-RC acknowledges that *autonomy* (see previous chapter) is a machine capability that requires particular characteristics in system controllers; humans are naturally capable of those characteristics, which already make them intelligent controllers.

STPA-RC has the following characteristics: 1) it builds on Thornberry's existing analysis and adds new parts, 2) it refines his original guidance, and 3) it introduces a method for identifying outside influences on the controller. The extended analysis is organized into eight parts, (a) through (h). The parts shaded in light blue in Figure 3-4 are portions of the analysis that apply in general, while the parts shaded in green only apply to humans. The analysis is meant to be straightforward and helpful, but it is not a perfect recipe for generating causal scenarios. Design and subject matter experts (SME) are still required to think critically during an STPA-RD analysis. The eight parts are not intended to be independent from one another, and there is no prescribed order to go through them during the analysis. In fact, hazard analysts often begin by examining the process model when performing Step 2. Regardless, the parts are discussed over the next sections of this chapter in the order of the parts shown in the figure.

Parts (a) through (e) of STPA-RC are maintained from Thornberry's analysis [190]. Although his controller model did not show it, his analysis included all the feedback (including human-sensory perceptions) being provided to the controller. This small step is meant to capture all the communications reaching the controller from a systems perspective, before the feedback is detected by the controller. In the context of this thesis, it provides traceability to the safety-control structure. In Figure 3-4, this is part (a), called *Information Availability*. It is neither light blue nor green because it examines general system information that is external to the controller. It acknowledges that feedback (FB) from lower entities, general communications (CC and IM), and control actions (CA) from higher entities are all inputs into the controller's available-information set.

Parts (b)–(e) are named after the four key areas of the OODA loop. Part (b), Observe, and part (e), Act, are the interfaces that a human controller has with the work ecology. Affordances are referenced in the action generation (e), and it is here that the analysis considers control actions to lower entities as well as feedback and communications to other entities. Additionally, a new concept is introduced called "affordance feedback" which is defined in the next section. Part (c), Orient, and part (d), Decide, apply to all controllers and examine the process model and control algorithm, respectively. Parts (f) through (h) are new in the extension and bring the extrinsic factors from Thornberry's human controller models (including physiology), as well as a new concept called "influences" (discussed in Section 3.3.7), directly into the analysis.

When Flach and Carroll discuss people in a complex sociotechnical system, they treat a human's behavior as a control problem containing distinct observation loops and decision loops [49]. The *observation* problem is signified in STPA-RC by the two-way relationships between Observe, Orient, and Decide in Figure 3-4. The human mind is always searching for information to update the mental model, clarify uncertainty, and compensate for noise and disturbances. Even after opportunities in the ecology are detected, the mental model must recognize information that it is searching for in the observed phenomena. The mental model itself is primed (by the nature of the decisions the controller is attempting to make) to search for and recognize particular observable features of the available information. The mental model informs those decisions, even if it has not been updated with the most recent available information. It is in these observation loops

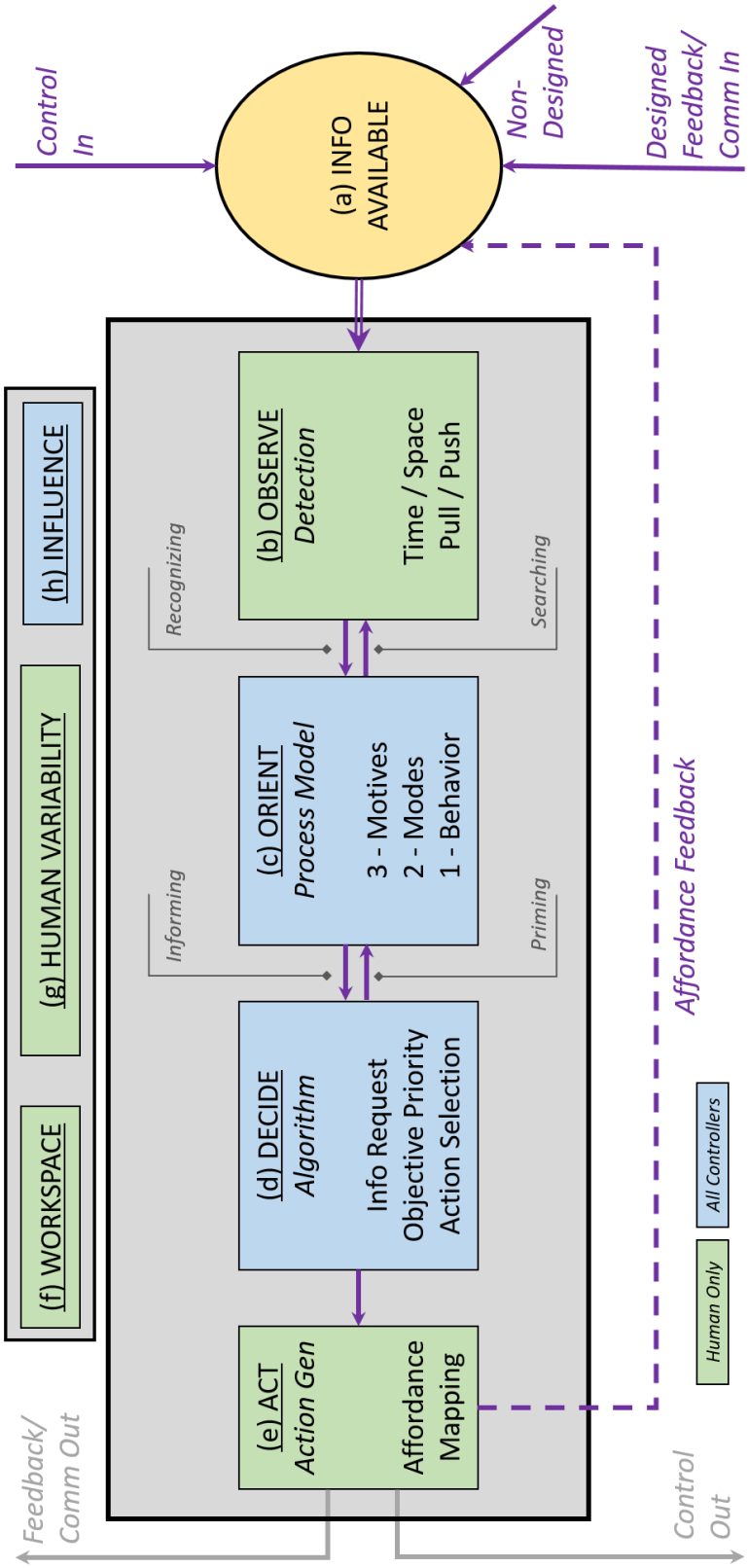



Figure 3-4. STPA-RC Analysis

Table 3-2. Comparison of Analysis Guidance (Stringfellow in Blue Text)

| 2014: Human Controller Causal Scenario Considerations | 2016: Refined Controller Causal Scenario Considerations |
|---|---|
| (1) Conflicting, incorrect, missing, delayed, or unrefreshed process states; no traceability to current affordance/actions | (a) INFORMATION AVAILABILITY<br>Control input missing, wrong, or conflicting<br>Designed feedback/communication missing, wrong, or conflicting<br>Non-design feedback/communication wrong or conflicting<br>Affordance feedback missing, wrong, or conflicting |
| (2) Process states undetected or interpreted incorrectly or too late | (b) OBSERVE<br>Controller does not fetch necessary display<br>Controller does not refresh expired display<br>Controller does not attend to appropriate display<br>Attended or exogenous display not noticed |
| (3) Process model inconsistent, incomplete, or incorrect | (c) ORIENT<br>Information not picked up<br>- Attended display misread or misinterpreted<br>- Accumulation of feedback not understood<br>Behavior model inconsistent, incomplete, or incorrect<br>Mode model inconsistent, incomplete, or incorrect<br>- *see mode-investigation table for details*<br>- Inadequate understanding of own control authority<br>- Lack of understanding of which processes are controllable<br>- Lack of understanding of process boundaries<br>Motives inconsistent, incomplete, or incorrect<br>- External motives conflict with system goals<br>- Inadequate mapping or resolution of high-level priorities |
| (4) Inadequate control algorithm | (d) DECIDE<br>Controller does not request or confirm required information<br>- Inadequate understanding of how to process feedback<br>Controller prioritizes wrong objective<br>- Local objective unknown<br>- Local objective violates safety constraints<br>Controller selects wrong action<br>- Did not learn control algorithm properly |
| (5) Action inappropriately afforded | (e) ACT<br>Action inappropriately afforded<br>- Controller does not understand how to execute control<br>Inadequate response mapping |

that many classic cognitive biases reside [3]. The *decision* problem is the one of classic control theory, in which a loop is closed between the controller and the controlled process, and errors in desired process behavior are measured from the feedback available to the controller (where it becomes an observation problem again). In the figure, the decision loop is not shown but is implied by the "control out" and "feedback in" arrows.

Table 3-2 shows a side-by-side comparison of Thornberry's analysis and STPA-RC. The new extension has updated parts (b) through (e) and incorporated applicable material from Stringfellow's taxonomy questions into the analysis, shown in blue text in the table. Extrinsic factor guidance—parts (f) through (h)—did not exist in the previous analyses, are not in the table, and is instead detailed further in Sections 3.3.6 and 3.3.7.

### 3.3.1 Information Availability

Part (a) of STPA-RC, the set of available information, corresponds to what Dekker calls "data availability" [4]. This section identifies the information presented to the controller

including controls, feedback, and communications. The information should be analyzed for the appropriateness of its content in light of what the controller needs to update its process model [38]. Part (a) does not consider the salience or formatting of the information; those aspects are considered elsewhere in STPA-RC.

This part of the analysis has been refined from Thornberry's version to explicitly differentiate between information available as originally *designed* to arrive at the controller and information outside original design intent (*non-designed*) that might be used by the controller regardless. Thornberry's analysis assumed that designed feedback is always in the form of intentionally planned displays, and non-designed feedback is only in the form of human-sensory perceptions. However, human-sensory perceptions could well be anticipated by designers, while there could be communication links, displays, or sources of information used by a human controller that the designer(s) did not anticipate. An example of this would be a copilot noticing how the pilot's hands are displacing a traditional yoke instead of looking at the copilot's own flight displays. The purpose of making the effort to delineate between designed and non-designed communications is crucial when new technologies and system upgrades threaten to change the nature of human-system interactions without properly documenting all the connections. The fact that pilots are using freely-available feedback (such as hand movement) that will disappear with a future upgrade (such as fly-by-wire[11]) is important for SMEs to realize. Another example would be the disappearance of traditional combustion-engine noise in newer, electrically-propelled road vehicles.

Those examples emphasize the importance of STPA Step 2. It is never enough for UCAs to be identified. Having discipline experts involved in a causal-scenario analysis is a valuable process that encourages the questioning of assumptions made about the design and operation of the system. Human-engineering experts who realize that pilots are using non-designed sources of feedback can inform the iterative systems-engineering process and have that type of feedback added to the control structure, where it is then considered designed feedback for future analyses.

Determining the appropriateness of the information provided to the controller is aided heavily by defining what states are required within the process or mental model [189]. During a system's concept development, or if the design is not yet mature, a detailed control structure may not exist, but hazard analysts can still use the process model as a guide. Human controllers must often interpret many different kinds of feedback to update the variables in their mental model. They use tiered communications, as discussed in the previous chapter [119]. The available information may even contain factors such as emotion or affect that cannot be analytically described [114].[12] Future research would benefit from developing methods to define and qualify these factors.

---

[11] In a fly-by-wire aircraft, a copilot's controls do not necessarily move when the pilot or flight computer controls the airplane, and the pilot also does not have to displace his controls by as large a motion as with a traditional configuration.

[12] Non-verbal human communication such as vocal inflection is an example. A former Air Force U-2 pilot described to me instances when landing his airplane in which the tone and urgency detected over the radio from the voice of his ground-based spotter were valuable indicators of his landing performance.

Thornberry suggested that the feedback the controller receives when an affordance is acted upon should be identified during the human-controller analysis. This is now explicitly incorporated in STPA-RC with *affordance feedback*. As discussed earlier, affordances mark the opportunities for humans to interact with the rest of the system and the larger work ecology. Affordances exist in physical controls and displays, as well as direct human-sensory perception of and action onto processes. Affordance feedback is based on the control-action generation, and it is defined for this thesis as *information that is received by a human indicating what has been commanded*.

Just because a human believes they have commanded an action does not mean the action has actually been received or executed. Affordance feedback is examined specifically to resolve that type of scenario. A *hard* affordance feedback is one received directly from the action interface. For example, a human would feel the tactile response of a spring-loaded button pressed with the finger to initiate a particular control.[13] A *soft* affordance feedback is artificially displayed to the human to indicate what was commanded to the lower process. Examples might be a light indicating that a valve has been *commanded* to open, or a *target* speed being displayed on a car's cruise-control display. Note that neither hard nor soft affordance feedback communicates the state of the controlled process, only information about the control request itself. Errors between commands and responses should be made clear.[14]

Documenting affordance feedback in part (a) is important for similar reasons as understanding non-designed information. For example, turning and removing a key from a traditional (non-electronic) car ignition is usually sufficient feedback for the driver to believe that the vehicle motor changed to a shutdown state; feedback from the controlled process (the car) was not required for the human to conclude the change had occurred. However, if the driver operates a newer car with an electronic key that does not rotate an ignition interlock and there is little or no sound from the electric motor, there could be cases where the driver removes the key and exits the vehicle without actually having turned the motor off.

The guidance included in Table 3-2 for part (a) is straightforward and meant to foster discussion on the correct timing, appropriateness, and nature of the feedback that a controller receives within the system. Sometimes displayed feedback to humans might not match directly perceived environmental phenomena. Spatial disorientation and vestibular illusions are a classic example of this [65]. Affordance feedback might not agree with controlled-process feedback, such as the electric-car example above, or designed feedback might not agree with non-designed feedback, etc., contributing to confusion, disorientation, or an incorrect process model.

---

[13] Another classic example is the ability for a pilot in an old airplane, in which the flight controls are reversibly linked to the aerodynamic surfaces, to feel stick forces when air pressure pushes back on the surfaces. During the historical evolution of flight controls, much effort was made in the handling-qualities engineering domain to mimic those stick forces even when the flight controls evolved to be no longer reversible [73].

[14] Leveson emphasizes the importance of measuring the "effect of the controller's action" [5, p. 296].

### 3.3.2 Detection

Any type of communication (i.e., CA, FB, CC/IM) might reach a human controller, yet there is still no guarantee that it will be detected or understood. Part (b) of STPA-RC, Observe, is a human-only portion of the analysis and corresponds to what Dekker calls "data observability" [4]. In Figure 3-4, part (b) is green because it only applies to human controllers. This section of the analysis examines *if* and *how* data are detectable and attended to in time and space. The distinction between data availability and data observability is important, because many scenarios and conditions could exist in which information is otherwise appropriate but not detected or comprehended by a human controller.

The temporality and spatiality of displayed and naturally perceptible information are analyzed here with some refinements from Thornberry's analysis. The guidance is updated and shown in Table 3-2. Once data are available, information can be derived from the interface or sensed environment by the controller or provided to the controller via the interface or sensed environment [59]. The controller can fetch available data that is not yet displayed, refresh an obsolete display, attend to an up-to-date display, or receive new data immediately via either the currently attended time-space or through exogenous cueing. Interface designers, human-performance engineers, and workspace-design experts may form an even more refined analysis based on the specific system. Many basic control and display principles begin to be applied here, such as the modalities (e.g., aural, visual) of the information presented to the human controller. SMEs may also apply principles of perception, such as static and dynamic presentations and psychophysical scaling of stimuli, as well.

An important consideration here is that it is useful to first identify the available information in part (a), as that frames the rest of the analysis on functional system relationships. This framework helps immensely with the process-model analysis, and it can guide the application of detailed engineering principles (e.g., human display-design standards) within the context of the system functions while preventing the tendency to apply those principles broadly as mere best practices.

### 3.3.3 Process Model

Boyd emphasized that *Orient* is the most vital area for analyzing behavior and decisions in a complex adaptive system [226]. This concept—part (c) of STPA-RC—is analogous to the STAMP process model or mental model. The analysis of the process model is so important it is often the first area visited in STPA Step 2. People are always updating their mental model with new expectations and observation strategies; they make associations after repeated stimuli and create templates of expected system behavior [145].

Before discussing the core set of guidance in this section, the relationship between parts (b) and (c) should be briefly mentioned. The two-way interchange between Observe and Orient, labeled as *searching* and *recognizing* in Figure 3-4, signifies one end of the intelligent controller's observation loop and corresponds to what Woods calls "information pickup" [227]. The first line in part (c) in Table 3-2 (information not picked up) reminds

the analyst to look at the information interface to determine its appropriateness for "representing the problem" [9, p. 16]. Whereas Thornberry included interpretation in his observe/detect section, in STPA-RC appropriate interpretation of feedback is emphasized in the process model section.

The remainder of the STPA-RC process-model guidance in Table 3-2 represents a significant update from the previous versions of the human analysis. Whereas previous versions of STPA simply ask if the process model is inconsistent, incomplete, or incorrect, STPA-RC refines this area to three types of system abstraction: behavior, mode, and motive. These levels are analogous to what Lee and See call "performance, process, and purpose" when they discuss the importance of agents within systems aligning their models to improve cooperation [114, p. 59]. In systems with controllers that adapt and reconfigure their priorities and responsibilities to handle a wide range of uncertainty, Ashby's law of requisite variety becomes important: as the variety of system capabilities increases, the available variety of information that controllers use must also increase [129].

*Behavior* represents how the controller's controlled process is performing and interacting with the mission environment. This corresponds to Leveson's model of the controlled process in Figure 3-2. It may be analyzed in the same manner as previous STAMP research has demonstrated, such as identifying different process model variables (e.g., airspeed, altitude) that the controller needs to know [189], [190]. Some process model variables are read directly from the feedback sensors, but some values may need to be calculated and stored as a different variable (e.g., altitude and airspeed represented as total aircraft energy). If the controller is human, the displays can aid in the process by dynamically calculating and presenting those translated variables. Analysts should document any translated variables to ensure they are appropriate for the tasks the controller needs to accomplish.

Some controllers understand more than just behavior states in their process model. Perhaps a controller might be sharing control of a process or controlling a lower component (or more than one component) which is in turn managing process behavior. In systems with many layers of controllers, another level of abstraction in the process-model analysis, *mode*, becomes important. A mode, as defined by Leveson, is a mutually exclusive set of system behaviors. There are three types of modes: supervisory mode, component operating mode, and system operating mode [38]. Table 3-3 defines the three modes and offers some refining questions to allow analysts to properly define the mode states that a controller must manage in its process model.

The *supervisory mode* captures the control relationships and communication links in the system hierarchy. When designing the supervisory relationships in systems, Leveson recommends the system incorporate redundant paths and allow incremental control of processes by humans [5]. If a higher controller is supervising a lower controller, it is important that the higher controller have current or immediate access to feedback about the process behavior that the lower controller is monitoring, should a contingency arise in which the higher controller must take control of the process. The supervisory mode can change during the operating process, affecting the priorities of controllers and communication paths. Stringfellow noted that human controllers maintain a model of not

Table 3-3. Mode Considerations

| | |
|---|---|
| **Supervisory Mode** | **The control relationships between the controller and component.** |
| | How is the controller operating over the component (e.g., direct, supervisory, sharing)? |
| | Which controlled components may apply authority limits and under what circumstances? Can those limits be overridden? How will conflicts be decided (i.e., who should have the final authority?) |
| | |
| **Component Operating Mode** | **The behavior of the controlled component.** |
| | What are the physical or logical assumptions and constraints associated with the component's current operating mode? |
| | What data in the information set is the controlled component using? |
| | What input/and output (interface) format is the controller using with the component(s)? |
| | |
| **System Operating Mode** | **The specified set of related behaviors of the system representing its operational state.** |
| | What operational state is the system in? |
| | Do all controllers know the current operational state? |

only the controlled process, but also of the organization [21]. A controller's awareness of the supervisory mode and supervisory structure includes considerations like which controllers in the hierarchy control which components, and which controllers share or hold priority over each. This is important when a controller must share workload with other controllers[15] to create behaviors that work in synergy with the behaviors of other components.

Also mentioned in the table are *authority limits*, which are a type of lockout or interlock that controlled components may exercise, by design, to ignore a received control request if it might be hazardous to the system. The controlled component might be designed to immediately ignore certain requests without further consideration, or it might first calculate that the request is hazardous and then ignore it. For example, a flight-control computer can limit the angle of attack the pilot demands, or a pilot can disregard an air-traffic control request if he sees visual traffic in the way. These authority limits must be carefully analyzed to make sure they do not prohibit behavior that might be necessary in some situations.[16] There must be some determination of who should have the final authority in case of a conflict. The concept of authority limits allows a lower controller in the hierarchy to perform or recommend actions against the wishes of a higher controller. This prevents overly complicating the safety-control structure or having to add or interchange

[15] Even more fundamental is ensuring that controllers know of the *existence* of other controllers that influence the components and processes it is trying to control.

[16] For example, a flight computer might incorrectly prevent human pilots from activating the reverse thrusters of an airplane after landing because the computer thinks it is airborne.

components, which takes away from the simple functional purpose of having the model in the first place.

The *component operating mode* is the behavior of the controlled component itself. An example would be an autopilot being in altitude-hold mode versus constant-velocity mode. Defining component operating modes includes identifying the environmental assumptions that must be met for the component to optimize behavior, the physical and logical constraints of the component mode and the feedback and communication information and data formats the component is using. To use modes effectively, the controller should have knowledge of the primary and alternate information sources for itself, its controlled components, and other controllers with which it interacts. Additionally, the input/output format the controller is using to control or monitor a component is important. A simple example of this would be a typist engaging a switch on a keyboard that converts keystrokes into function commands instead of raw-letter inputs.

The *system operating mode* specifies sets of related behaviors of the system representing its operational state. It is important that when a subphase of operation (e.g., taxi, takeoff, transition, and landing/recovery) changes that it is communicated to all controllers, because the modes might default to different settings. More broadly, there should be clear communication between controllers on all pertinent system modes, supervisory structures, and subphases [174]. Fleming proposed that a meta-controller might be necessary sometimes to assign control modes and mode changes within the system. This meta-controller could be a real-time entity in the control hierarchy and/or simply a set of policies and procedures that the controllers within the system follow to change modes and communicate those changes when appropriate [185]. An example of this is an ambulance siren communicating a temporary emergency to everyone on the road. Drivers, because of training and policy, change their observation patterns and control algorithms to be able to move out of the way for any public-safety vehicles until the emergency has passed.

The mode and phase considerations in the STPA-RC process-model analysis would be part of both Leveson's model of the automation and model of the context in Figure 3-2. The controller's component-operating modes are a property of its controlled components, while the supervisory system operating modes are properties of the entire system. Analysts may use guidance on identifying and preventing mode confusion such as those proposed by Leveson and discussed in the previous chapter [38], [91]. Other design considerations include providing controllers a thorough understanding of how controlled components behave in off-nominal situations[17] and providing salient feedback to the controller about modes and phase transitions during these situations.

*Motives* are the third level of abstraction in the process-model analysis. Whereas behavior and modes describe hard-systems properties—the syntax (e.g., physics and dynamics)—motives address the soft-systems properties—meanings and semantics [62]. As Checkland states, "Any purposeful human activity implies commitment to a particular

[17] An example would be a car driver knowing that the cruise control will disengage if the vehicle goes up a steep hill.

set of values" [45, p. 126]. Motives should not be confused with local objectives. Any controller, through the control hierarchy, can be given any number of cost functions to maximize by optimizing the behavior of its controlled process(es). It may also choose different modes and communicate with other controllers to optimize various behaviors within the system as uncertainties in the environment arise. However, the system itself exists to achieve "good enough" results [45, p. A49]. The intelligent controller's understanding of these motives—of which there can be many (and sometimes they may conflict)—also affects the decisions it makes.

There is no corresponding model for motives in Figure 3-2. STPA-RC makes the first effort to capture the controller's managing of motives explicitly in the analysis. Local objectives have always been considered a part of the control algorithm, and that is still an appropriate place. However, the motive(s) can affect local objectives. Two aspects of motives are important here. The first is *external motives*, which is an understanding of any motives the controller personally maintains outside the designed system. An example would be the personal pressure to get home early that might prompt a pilot to ignore intentional safety constraints and rush a landing. Humans always bring with them unique sets of motives into their systems, and those may be affected by the system itself or other humans in it; it is important that appropriate motive-based priorities (e.g., safety, performance, schedule, personal well-being, etc.) are consistently understood and prioritized by the controllers in the system.

The second aspect is *motive mapping*, which is the controller's understanding of how motives at top levels of the system map to objectives at the controller's level. There are several methodologies in the literature that attempt to formalize such mappings. For example, Rasmussen introduced the means-ends hierarchy that was discussed in the previous chapter [228]. A future research effort should study these methodologies and similar approaches to further advance the capability for STPA to examine a controller's motive model. Care should be taken to ensure that not too much flexibility is permitted in the system. Human controllers will exhibit exploratory behavior within their spheres of control as they constantly try to optimize behaviors to prioritize and trade off between the important system motives [5]. This behavior is not necessarily intended to violate safety, but that result could occur if there is no consistent feedback that an unsafe condition is possible or imminent [229]. Feedback that highlights violations of important safety constraints should be provided, and if so the controllers should understand why those constraints are important [9].

Table 3-2 guides the causal-scenario analysis by emphasizing that the process model can be inconsistent, incomplete, or incorrect, and it refines that analysis into behavior, modes, and motives. The human mental model is always changing in both structure and content [190]. Accounting for those three types of system abstraction allows human controllers to be adaptive. What is important about the process-model analysis is that it be used to identify potential discrepancies between the process model and the real states of the system. If operators have begun using a differently structured process model of the system than the designers intended, updates to the design should provide the appropriate communications so that those models are appropriately supported. During STPA Step 2, human and software engineers should be encouraged to work together to

identify causal scenarios that develop due to the manner in which computers and people share information and update their process models.

### 3.3.4 Control Algorithm

The transition in STPA-RC between the process model and the control algorithm—parts (c) and (d), respectively, in Figure 3-4—covers another two-way interchange. It is labeled as *priming* and *informing*, and it signifies the other end of the intelligent controller's observation loop, this time occurring between Orient and Decide [49]. The full observation problem encompasses Observe, Orient, and Decide, and shares similarities with Endsley's concept of situation awareness (SA): perception of the elements, comprehension of their meaning, and projection of their status into the future [110]. SA is not only the process model, nor is it something that a controller simply gains or loses. The maintenance of the variables in the process model is a continuous activity that always exists across observations, schema-building, and projecting. Humans are always updating their schemata, and erring is learning [150].

The decision problem, that of classic control theory, is based on the controller applying the appropriate actions based on the information in the process model, and receiving feedback after those actions have propagated through the controlled process or component. STPA-RC makes no recommendations for which specific decision-making and schema-building theories should be applied to parts (b) through (d). Some prevalent theories in the literature were presented in the previous chapter. Table 3-2 establishes a simple three-part guidance for part (d): required information not requested, wrong objective prioritized, and wrong action selected. The appropriate considerations from Stringfellow's taxonomy are also added in blue text to that part of the table [21].

There could be many explanations for an inadequate control algorithm, including cognitive biases, and only designers, engineering SMEs, and expert practitioners can truly elucidate the scenarios that include inappropriate decision-making. One type of contributing factor to inadequate control is the existence of wrong procedures or conflicting procedures, such as instruction manuals or policies. Section 3.3.7 introduces some additional guidance in STPA-RC for analysts searching for such scenarios.

Controllers might make use of decision aids to supplement their process model and control algorithm. These aids could come in the form of information translators, as mentioned in the previous section, that dynamically convert feedback into a format that is more easily usable for understanding the process and/or system and making appropriate decisions. Another type of dynamic aid goes one step further by suggesting one or several actions that would appropriately optimize the process being controlled.[18] There are also static aids, in the form of checklists, procedure references, guides, and information sheets that can be referenced by controllers in real time.

[18] An example would be the flight-director feature in some modern airplane cockpits. Here, a continuously-updated moving symbol is displayed to the pilot indicating where to point the nose of the airplane in order to achieve a desired navigational path.

This thesis does not recommend any specific method for modeling decision aids in the hierarchical control structure. That area would benefit from future research. The ability to understand and prioritize objectives and decision options is important for efficient control, but efforts must be taken to provide human controllers with an appropriate amount of assistance [53]. STPA-RC simply acknowledges that finding and translating appropriate feedback, determining objective prioritizations, and even choosing an action may be aided by software. Furthermore, as mentioned in the modes discussion, controlled components may go beyond suggesting appropriate actions and actually limit the controller's authority. A thorough documentation of the supervisory structure is thus important for capturing the relationships between controllers to support the analysis.

### 3.3.5 Action Generation

Part (e) of STPA-RC is the generation of a selected action. It is once again an analysis of the interface between a human and the work ecology, and in Figure 3-4 part (e) is green because it only applies to human controllers. Thornberry's guidance is included here, along with one of Stringfellow's considerations in blue text (controller does not understand how to execute control) [21]. One refinement has been made here by adding a *response-mapping* consideration. Many human-engineering principles exist in workspace design to address stimulus-response compatibility [59].

It should be emphasized that correct (or satisfactory) action *selection* is a function of the control algorithm. Executing the action, whether through generation of a motor movement or any other form of information output, is what part (e) examines. Some decision-making theories, such as Rasmussen's S-R-K framework or Kahneman's Type I/II framework argue that some reactions are directly executed based on the process model without being selected through a control algorithm, making those types of responses more reflexive in nature [145], [165]. Regardless, it is the proper execution of actions (whether they are purposefully or instinctively selected) that is of importance during the analysis here.

### 3.3.6 Extrinsic Factors: Human Engineering Considerations

Leveson and Thornberry included extrinsic factors in their human-controller models; STPA-RC presents parts (f) through (g) in Figure 3-4 as new explicitly-defined sections of the analysis to examine those factors. Extrinsic factors contribute to scenarios grounded in the intrinsic mechanisms of the controller—those included in parts (b) through (e) of STPA-RC. The new portions of the analysis provide more refined guidance to help generate scenarios that might have been missed looking at only the OODA portions of STPA-RC. The first two extrinsic factors, (f) and (g), are discussed in this section. They are the human-engineering considerations of workspace and variability, colored green in the figure.

There exists a wealth of standards, guidance, and best practices in various industries for complementing human characteristics with system work environments, such as MIL-STD-1472 and related documents [28], [41], [42], [121]–[128]. STPA-RC does not recommend any particularly detailed direction, but instead borrows from common human-

engineering themes to provide analysts with some basic guidance to supplement the expertise that human-engineering SMEs can provide later if needed. The important feature of STPA is that it targets a specific system's design and use philosophy. It *first* identifies hazardous behavior (based on top-level requirements and the control structure), *then* it identifies the causal scenarios for the specific system behaviors already identified; it is here that human-engineering standards and design guidelines can inform the analysis, once the safety context and constraints are established by the earlier portions of STPA.

*Workspace* considerations (f) cover the human controller's relationship with the task setting and interfaces. The following considerations should be included in the analysis:

PART (F) – WORKSPACE

- Climate
  - Lighting
  - Temperature and pressure
  - Audial environment
  - Other physiological phenomena (e.g., inertia, vibrations)[19]
- Physical Ergonomics
  - Anthropometric constraints (all genders)
  - Kinetic constraints (all genders)
- Task Workload[20]
  - Persistence of interactions[21]
  - Multiple-control responsibilities[22]

*Human variability* considerations (g) cover characteristics that fluctuate between different people. These considerations are important because designers, practitioners, and safety engineers must work closely to determine the balance between selecting a particular subset of the general human population to be in the system and designing the system to accept a wider range of human traits. If STPA is being conducted on an already-existing system with a known user pool, these considerations can be used to determine the adequacy of the system given the existing range of characteristics of its human controllers:

PART (G) – HUMAN VARIABILITY

---

[19] The aerospace industry, particularly in high-performance aircraft and spacecraft, must account for this.

[20] A thorough review of guidelines on designing for appropriate workload is not included, as the intent of STPA-RC is to provide only basic guidance about human-engineering factors.

[21] Persistent interaction is associated with decreased vigilance, alarm fatigue, and change blindness [104], [117].

[22] Simultaneous control of multiple dissimilar activities (e.g., manual control of a process combined with verbal supervision of a controlled component) by a single human should be carefully analyzed in the context of the work being accomplished [101].

- Physical attributes
  - Perceptual acuity (visual, audial, etc.)
  - Athletic and motor acuity
- Mental attributes
  - Attention capacity
  - Psychological and emotional health
  - Risk tolerance and proclivity to trust
- Health
  - Age
  - Fitness and nutrition
  - Injuries, diseases, and disabilities
  - Drugs and medications
- Physiological stress
  - Fatigue
  - Sleep and shift cycles

This guidance in STPA-RC allows classic human-engineering principles to be included in the causal-scenario analysis. Many of the considerations in parts (f) and (g) may contribute to causal scenarios jointly (e.g., stress and workload). The purpose of listing extrinsic factors is, like the other parts of the analysis, not to provide a set of guidelines that are independent or exclusive from one another, but to assist with the analysis of the inherent behavior of the controller.

### 3.3.7 Extrinsic Factor: Influences

Stringfellow presented a list of guidance to support her error taxonomy, including items like interfaces, human cognition, and physiology, which have been incorporated into STPA-RC as discussed in the previous section. The remainder of her guidance includes items like experience, resources, training, culture, and procedures [21]. These considerations are incorporated into part (h), called “influence”. This part is different than the other extrinsic factors in two ways. First, it is applicable to any controller—not just humans—and is thus light blue in Figure 3-4. Second, it requires thorough knowledge of the entire hierarchical control structure and current operating philosophies of the system, which in turn absolutely necessitates that expert practitioners, not just engineering SMEs, assist with this part of the analysis.

Flach and Carroll emphasize that in a sociotechnical hierarchy there are controls and communications between the many levels, some more informal than others. They describe organizations as having “interacting, nested closed-loop dynamics that span the multiple social layers,” and suggest that “every element in [the] system can be influenced

by output from every other element" [49, p. 6]. Leveson calls this a "complex causal network of relationships" [9, p. 18]. A proper STAMP control structure captures all the connections in a system, and as discussed earlier, STPA-RC includes methods to iteratively update the structure as non-designed connections are discovered.

A hazard analysis of a complete system should include the control actions of all controllers in the organizational hierarchy. However, some analyses might choose to focus on a single phase of work, such as the operating process. In the example of air transport, the safety control structure used for the analysis might have the air-traffic controller or pilot as the highest entity, and then all the other entities and control relationships of the operating process would be included, such as the flight computer and airframe. When STPA is performed on a control structure that only models the operating process, without including the higher levels of organizational hierarchy, controls from these higher levels should not be ignored; the phase being examined in the hazard analysis is still affected by the rest of the system.

While the previously discussed workspace and variability guidance in the hazard analysis seeks to identify how a human controller is affected by human-engineering design and inter-controller diversities, respectively, the guidance presented here seeks to identify how a controller is affected by information and actions that come into existence at some point in the system and continue to exist. These *influences* are controls and communications that would be included in the STPA analysis if it were performed on a larger safety control structure that included the entire organization. However, in a hazard analysis performed only on a phase, part (h) of STPA-RC may be used to fold the higher sociotechnical hierarchy into considerations that shape a controller's real-time behavior during the phase. Hajdukiewicz et al. differentiate between influences and physical constraints (natural laws) [170]. For example, two airplanes flying in the national airspace do not avoid colliding with each other because it is physically impossible, or because there is a passive failsafe designed into the air-traffic system. Rather, there are influences in the form of right-of-way rules, social contracts, and procedures that shape the real-time air-traffic control and piloting actions that prevent vehicle-separation hazards. Influences can be viewed as controls and constraints on behavior that evolve outside the time scale of the phase.

Figure 3-5 emphasizes the focus of part (h). In this figure the phase chosen is the operating process, but it could also be a maintenance or planning process, for example. For the remainder of this discussion an operating process is used. The arrow pointing down (denoted as "control") signifies the controls from the higher levels of the organization that are not captured by the safety control structure of the operating process. Because those controls and communications typically operate at larger time constants, they evolve over a period of time that includes, but also precedes, the active time of the operating process. Entities within the operating process are still affected by these influences. A simple example is the training received by operators in an organization. Additionally, actions performed in other phases can also affect the operating process. The horizontal arrow (denoted as "process input") signifies this with a type of influence called a *setting and/or configuration*. The previously discussed examples of the maintenance technician making an inappropriate adjustment to a system component or the worker that forgets to reset his

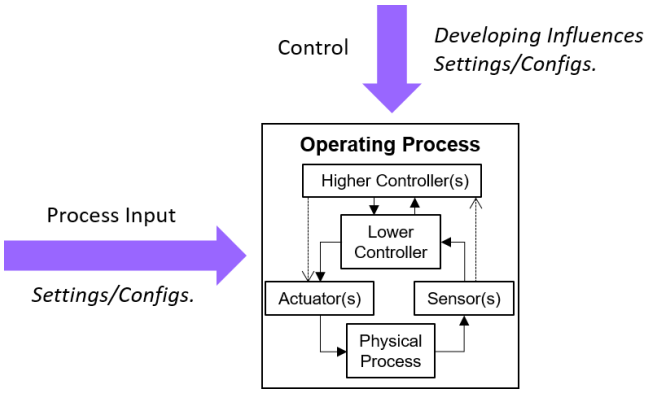



Figure 3-5. Organizational Influences

display settings before handing operations off to the next worker apply here. Settings and configurations can also be controls from higher in the organization. An example would be someone in airline management sending a memorandum to all pilots to temporarily stop using their flight management software to aid in descents.

A diagram to incorporate these concepts into STPA-RC is presented in Figure 3-6, which complements Figure 3-5 and introduces the guidance for part (h) of the analysis. Each block in the figure represents a type of influence that generally begins evolving at some point and continues to affect entities during the operating process. The chosen categorical divisions between blocks (influence types) are motivated by Schein's levels of organizational culture, but for STPA-RC the divisions should be considered arbitrary [43]. Influences can affect both human and non-human controllers of the operating process, and the figure includes the same color scheme (i.e., green for human controllers, light blue for any controller) as Figure 3-4. Although the ordering of the blocks might sometimes reflect the relative length of time it takes each type of influence to evolve, the blocks should not be interpreted as a chain of events.[23]

The first four types of influences—societal culture, organizational culture, behavioral standards, and rules and techniques—are called *developing* influences because they typically develop and change with larger time constants as compared to the time constant of the operating process. Although developing influences are simply controls in the socio-organizational hierarchy, if the hazard analysis is being performed only on the operating process, the causal scenarios having to do with culture, standards, rules, etc. would be identified by examining them as influences during STPA-RC. It should be noted that societal culture, organizational culture, and behavioral standards affect only humans, while rules and techniques can also affect non-humans.

[23] The intent is not to find an aspect of societal culture that causes an aspect of organizational culture that in turn causes an aspect of behavioral standards, and so on. Aspects from each category may be identified individually, and each aspect can directly influence controller behavior during the operating process.

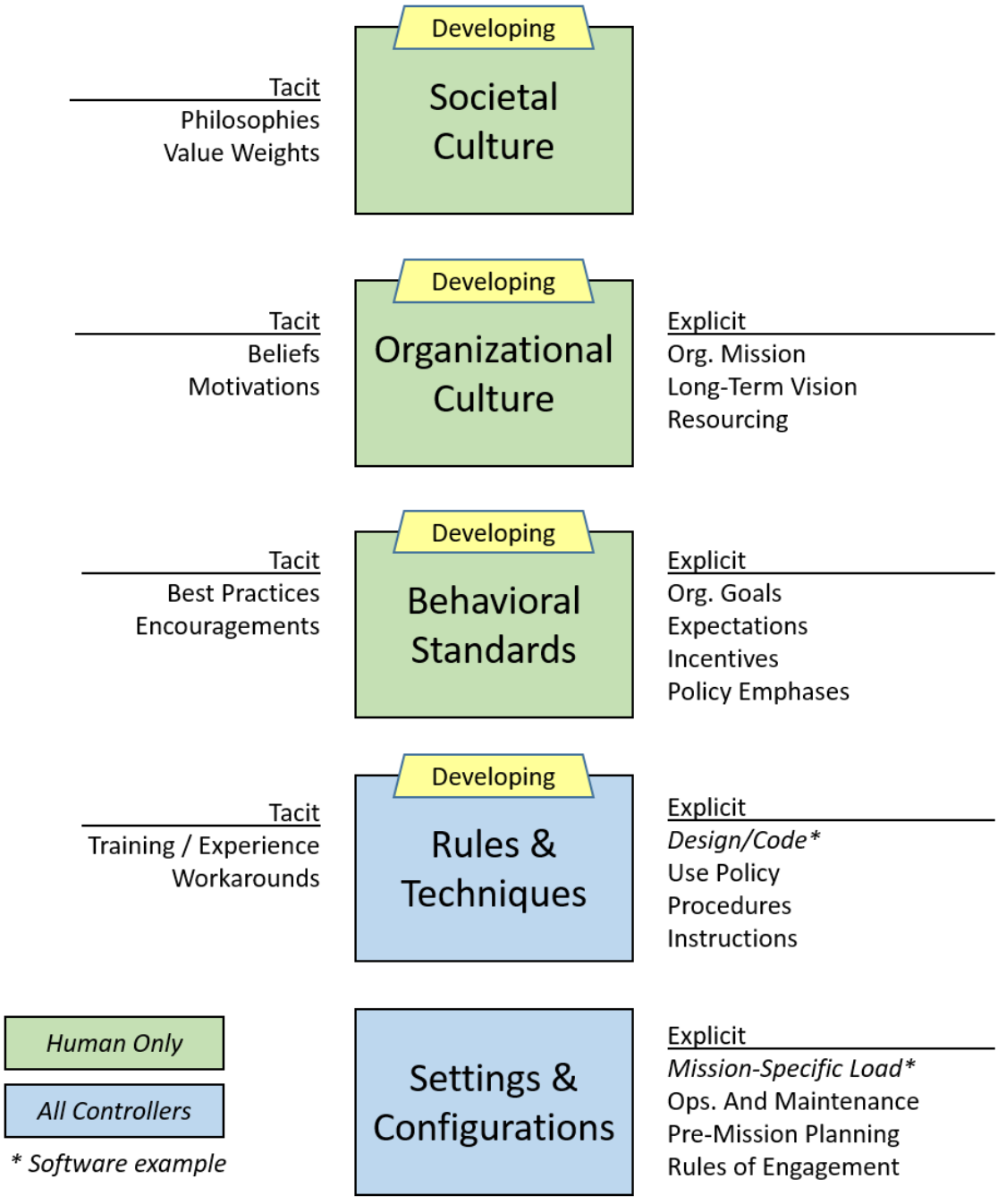



Figure 3-6. Types of Influences

Next to each block in the figure are guidelines for both *explicit* (formal, articulated, and codified information) and *tacit* (information learned by association and not easily transferred via media) influences. In an example of an explicit influence, a company policy letter would mandate employees discontinue use of a certain feature of an assembly line's computer display based on data gathered from a formal employee reporting system. On the other hand, a tacit example would be assembly-line operators on their lunch break complaining about that feature of the computer, and those sentiments eventually evolving into a proclivity not to use it. Both of these cases describe "rules and/or techniques", which are discussed more below.

A non-human controller is affected only by explicit influences, while humans are also affected by tacit influences. Experience and learning are integrated over time, shaping the human's decision-making algorithm and schemata. Explicit sources might be the only ones easily discoverable by a hazard analyst with little experience in the specific domain; this is the underlying reason that expert practitioners are crucial for part (h), as they understand sources of tacit knowledge in the organization.

*Societal culture* refers to norms and values inherited from a person's societal identity, beyond his professional organization. Societal culture can only be tacit because a person receives it throughout his life, via his geographical upbringing and social experiences. Safety attitudes, for example, can be largely swayed by nationalistic mindsets [230]. Additionally, cultural characteristics might dictate how humans approach decisions.[24] *Organizational culture* rests in the sociotechnical organization's identity and beliefs, and influences many aspects of the innate desires of controllers to achieve the purpose of the organization (or even the industry as a whole). These influences can begin to take on explicit form through leadership's mission and vision statements, and long-term resource planning devoted to the big picture can affect the safety attitudes and capabilities of practitioners in the organization [232].

*Behavioral standards* cover the manners in which designers and practitioners communicate, cooperate, problem solve, and trust each other to accomplish the goals of the organization. Here, specific criteria and expectations for behavior are enforced, and incentives (i.e., rewards or punishments) and emphases are established explicitly to maintain a consistency of controller capability. An example would be the aerospace industry's crew resource management (CRM) standards, which dictate how pilots in a cockpit should share responsibilities and communicate effectively for normal and abnormal situations [233]. Standards may be established by the organization, or by lower units within the organization. Behavioral standards do not have to be design-specific and often are not. Tacitly, practitioners will also develop their own best practices and adjust their standards of behavior based on how they perceive (actual or rumored) other people and methods of behavior to succeed or fail. Organizations should maintain information repositories that attempt to make explicit many of these otherwise unwritten standards. A good example is

[24] Members of different cultures might, for example, expect a control lever to move in different directions in order to achieve an identical function. More generally, there is possibly variability among different world populations in areas like "fairness, cooperation, spatial reasoning, categorization and inferential induction, moral reasoning, reasoning styles, self- concepts and related motivations" [231, p. 2].

the AF Military Flight Operations Quality Assurance (MFOQA) Program, discussed more in the next chapter [234].

*Rules and techniques* cover specific procedures and mechanizations for operating (or preparing/maintaining) the system given its specific design and mission. When the controller is a machine or software, rules and techniques exist in the physical design and/or software code. When the controller is human, explicit influences include written use policy, operating procedures, and system manuals and instructions. It is also not unusual for humans to tacitly develop techniques and workarounds to make their work easier, more efficient, or to compensate for perceived mistakes in design. Efforts should be made to minimize discord among explicit rules and techniques and also between explicit and tacit rules and techniques.

A common type of tacit rules-and-techniques influence is the regular training and practice that human controllers undergo to gain knowledge and experience on the system. Learning and repetition encourage assimilation and accommodation in the controller's mental schemata [147]. Experience and expertise development improve motor memory, mental-model efficiency, and decision-making abilities [59], [165]. Regular proficiency exercises are required to reinforce these abilities, and operators should also practice and retain skills required for degraded modes of system operation [215]. Similarly, crucial emergency techniques need to be reinforced regularly.[25] Furthermore, proficiency with other agents in the system (human or controller) allow for the human to develop trust, which can only happen over time after the person has perceived an alignment of his behavior, modes, and motives with the other agents' [114].

Settings and configurations are any activities, controls, or communications that specifically affect a single operation.[26] These types of influences can come from the organizational hierarchy above the phase (control), in the form of provisional rules of engagement, one-time objectives, or modified organizational incentive structures.[27] Settings and configurations can also come from the phase preceding the operating phase (process input). Previous operations and maintenance (O&M) phases, for example, might have an effect on how the system is configured before it commences its process. An inspection that does not notice fatigue in a wing spar or a cockpit switch left in an unexpected position can affect the dynamics of the following flight of an airplane.

Ideally, influences are all intentional and aligned across culture, standards, and techniques [236]. However, there are always unintentional influences in human-activity systems, both explicit and tacit. An example might be conflicting policies or outdated

---

[25] The military uses a training technique called *boldface*, in which operators memorize and are regularly tested on itemized lists of emergency procedures that cover critical situations, such as a rifle jamming on the battlefield or an aircraft engine catching fire in flight [235].

[26] Settings and configurations can also take the form of temporary rules or constraints that might affect several operating phases, but are subject to being rescinded or adjusted at any time.

[27] For example, an airline may assign its pilot fleet a particular priority stack (e.g., fuel efficiency more important than flight time) for a particular day of operations. Pilots would be briefed about this before flying, and the navigation software would be configured as such on each airplane with a day-specific load. Another example is mangers of the national airspace system (NAS) maintaining a database of temporary notices for pilots to read every day.

procedures resulting in the wrong control algorithm making its way into the operating phase, or employees perceiving a supervisor as implicitly rewarding or punishing certain types of behavior and slowly assuming different standards because of it. One of the goals of part (h) of STPA-RC is to identify unintentional influences and ensure that they do not interfere with intentional constraints or unnecessarily limit naturally-possible behavior that might be required for safety. This insight allows analysts to find causal scenarios that involve controllers executing inappropriate algorithms because of poorly standardized influences.

#### 3.3.7.1 Policy Mapping

Policies within a sociotechnical system are one method in which explicit influences can be established. Policies can originate at any level in an organization, and they can document a range of explicit influences including organizational culture, behavioral standards, rules and techniques, and even settings and configurations. The level of organization the policy stems from and the type(s) of influence the policy documents are independent.

Policies in a system can be reviewed to understand the way the system is intended to function. This includes the explicit visions, goals, standards, and instructions that cover the development, testing, and fielding of systems and products. Many organizations also document their safety programs and practices. The members of an organization can benefit from being aware of the number of published policies in their enterprises, the scope of each, and relationships between these items. The ability for employees to understand and shape policy is crucial for promoting a healthy safety culture [4]. Having clear and accessible insight into policy items allows the following to be considered:

- Are there gaps or holes?
- Are there redundancies or inefficiencies?
- Do any items conflict?
- Is the manner in which items trace through the organization and reference other items clear?
- Are items readily accessible, consistent in format, and updatable?

An organizational policy review (along with organization charts and documented activates) can assist hazard analysts to model the safety-control structure. This includes identifying control and feedback relationships, determining levels of authority, and characterizing the phases of work. Policy can also inform the allowable modes of the system and its components, including the subphases of operation.

Because influences are not the same as physical constraints, policy does not bound possible system behavior. In fact, a human controller might violate a published procedure if they perceive an immediate threat to a motive they deem important (such as safety). Instead of visualizing controllers as being bounded by policies, it is better to consider the policies as internalized (and interpreted) by the controllers. Policies, like all influences, affect the decision-making algorithms and schemata of the human controllers in the operating process. Within the operations, the possibilities of a controller’s behavior are

determined by their control algorithm and mental model while bounded by physical constraints.

A method to visualize the policy in an organization can be useful for the reasons mentioned above, including traceability, accessibility, and maintenance of consistency. For this research, a new visual aid was developed called an Explicit-Influence Map (EIM). It is a common planning tool—like a STAMP control structure—that allows members of the organization to see how explicit influences (including policies) trace down to front-line practitioners. EIMs may take unique forms depending on which stage of product development is of concern. A tester, for example, is affected by different local policies than a field user, although their EIMs might begin to look similar at higher levels of organizational policy.

An EIM generated for the AF test enterprise is presented and discussed in Appendix B. The next chapter discusses the organization and safety programs of AF product-development and references the EIM as an aid to understanding the applicable policies therein. Beyond being a useful tool for research, the stakeholders of the organization itself can also use the EIM to visualize their own policy domain, quickly reference information, and shape the influences that affect their practitioners.

## 3.4 Example: In-Trail Procedure

A new type of air-traffic procedure called Airborne Traffic Situation Awareness In-Trail Procedure (ITP) has been proposed as part of the Federal Aviation Administration (FAA) Next-Generation air traffic control (ATC) modernization. The FAA established ITP's initial safety and interoperability requirements in 2008 [237]. That analysis was based on Fault Tree Analyses (FTA); in 2012 a new report by Fleming et al. performed an STPA analysis of ITP [238]. A detailed description of the ITP concept and the comparison between STPA and traditional findings are not covered here; however, ITP is explained as appropriate to describe the refinements in the analysis of human controllers using STPA-RC.

In 2014, Thornberry took one example from the STPA report (a flight crew executing one of the ITP maneuvers) and applied his human-controller analysis (modeled in Figure 3-3) [190]. As has been shown in Table 3-2, Thornberry added guidance for data observability and action affordance to the STPA human analysis, and he identified new causal scenarios for the ITP scenario based on those additional considerations. The remainder of this section again uses the ITP example and apply STPA-RC to refine the causal scenarios for the human controller.

ITP enables more flight-level (FL)[28] altitude changes for aircraft in transoceanic airspace in order to improve flight efficiency. Currently, transoceanic flight is performed

[28] A flight level is a type of altitude reading that is calculated by using a standard barometric sensor setting of 29.92 inches of mercury for all aircraft sharing an airspace that uses the FL altitude-assignment procedure. It is called a pressure altitude, and it is rarely the true altitude above sea level because barometric pressures

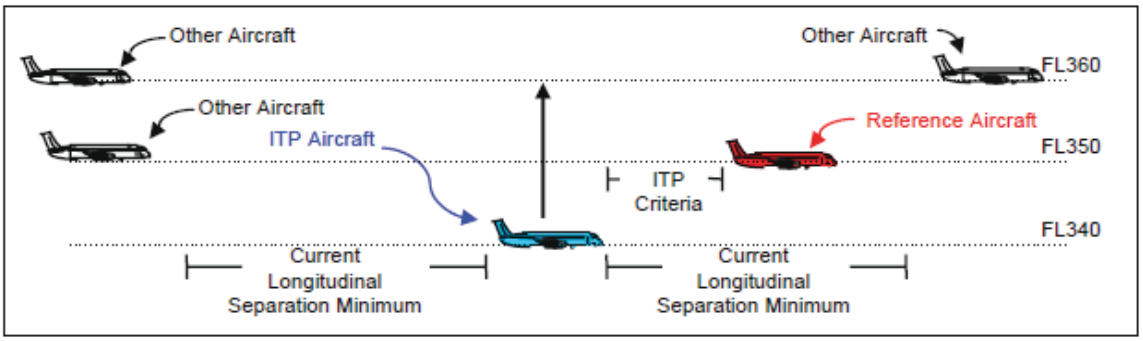



Figure 3-7. ITP Following Climb [238]

under procedural control. Aircraft traverse the oceans via predetermined flight paths and altitudes, parts of which cannot be monitored by ATC radar. Aircraft enter these paths and then make voice-radio position reports to controllers until they exit on the other side. Because there is no positive radar control, altitude changes are only granted by ATC when very conservative distance-separation minima between aircraft are met based on the verbal position reports.

The Next-Generation concept involves additional equipment installed on modern aircraft. One set of equipment is called Automatic Dependent Surveillance-Broadcast (ADS-B), which allows aircraft to share their navigational data—including position, velocity, and altitude—with any other aircraft that is equipped to receive the appropriate transmissions. For transoceanic flights, it is assumed that ATC does not receive these broadcasts for the purpose of this example. Another set of equipment is simply called ITP equipment, which is an upgrade or modification to an aircraft's flight computer that allows it to determine, onboard, whether criteria such as the distance and Mach (speed) differential between aircraft are suitable for an FL change. The full list of ITP criteria is available in the original document [237].

Figure 3-7 displays the example that Thornberry chose from the original STPA report [238]. The ITP maneuver for the example is called a following climb. In it, the aircraft of interest (in blue) is flying at FL340 and wishes to change to FL360 as indicated by the arrow pointing up. The current procedural-control distance-separation minima are shown as brackets in front of and aft of the aircraft. Using current procedures, the aircraft of interest would not be given permission to climb because it would violate separation minima with the reference aircraft (in red). The ITP distance-separation minima, however, are lower than the current minima based on the other ITP criteria being met as calculated by the ITP equipment. The narrower ITP criteria would allow the blue aircraft to initiate the FL change. To initiate an ITP following climb, the flight crew would: determine the ITP criteria are met from their flight computer; verbally request the ITP maneuver from ATC; verbally transcribe the ITP criteria to ATC; receive clearance from ATC to execute

---

are often not exactly 29.92 inches of mercury. However, atmospheric pressures vary dynamically across large distances, so if all aircraft use the same sensor setting they will read the same altitudes on their instruments, making vertical deconfliction possible. "FL300" corresponds to a pressure-altitude reading of 30,000 feet.

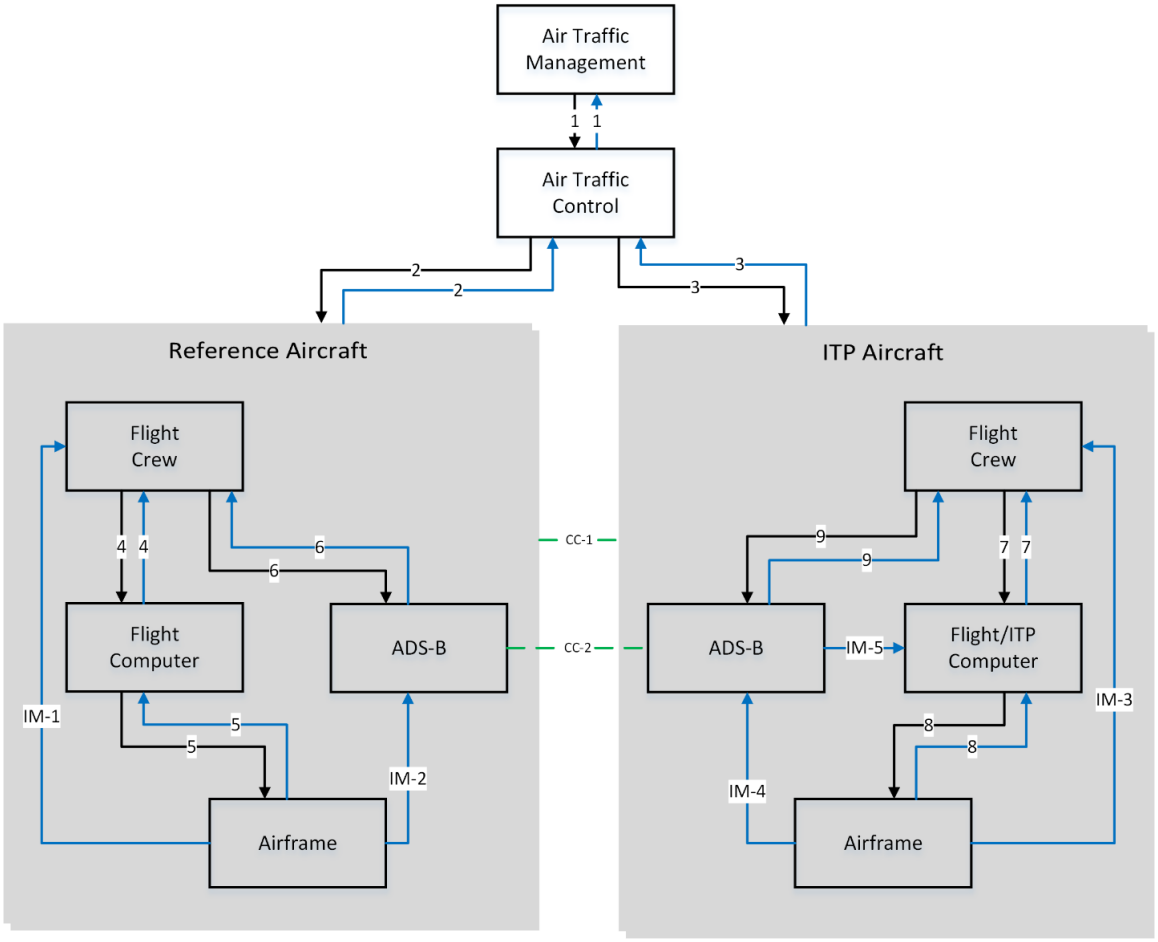



Figure 3-8. Safety Control Structure for In-Trail Procedure

the climb; then before executing, confirm that the airspace is clear of other aircraft and weather using visual feedback as well as indications from the ADS-B and the aircraft weather radar.

A top-down safety approach begins with the identification of accidents and hazards. The original STPA report considered one accident, human death or injury, and identified the following hazards related to it [238]:

H1: A pair of controlled aircraft violate minimum separation standards

H2: Aircraft enters unsafe atmospheric region

H3: Aircraft enters uncontrolled state

H4: Aircraft enters unsafe attitude (excessive turbulence or pitch/roll/yaw that causes passenger injury but not necessarily aircraft loss)

The ITP safety-control structure is shown in Figure 3-8, which has been modified from the diagram in the STPA report to use my visual-formatting style. The ITP aircraft is on the right and contains ITP equipment in its flight computer. Both it and the reference aircraft each contain ADS-B equipment that uses feedback from each airframe (IM-2, IM-4) to

Table 3-4. ITP Variable Reference (Abbreviated)

| Variable | Name |
|---|---|
| Control 7a | Initiate ITP |
| Control 7b | *etc.* |
| Feedback 7a | ITP Criteria |
| Feedback 7b | Weather-Radar Data |
| Feedback 9a | Nav Data (All) |
| Indirect Measure 3a | Motion |
| Indirect Measure 4a | Nav Data (Self) |
| Comm 1a | Sight of Other Ship |
| Comm 2a | Nav Data (Other ship) |

calculate and broadcast navigational data (CC-2). ADS-B data are displayed to each flight crew (FB-6, FB-9), and the ITP equipment also receives ADS-B data (IM-5). Both flight crews may communicate with ATC via voice radio (2, 3). If the aircraft are within close range of each other, the flight crews may see each other's aircraft visually (CC-1). The status of the ITP criteria and weather radar are displayed to the ITP flight crew (FB-7).

Depending on the system modes, there may be different possible communication protocols, individual responsibilities, display symbologies, environmental assumptions, and performance capabilities among the various components and processes in the control structure. As the ITP concept transitions to a detailed design, the types of possible modes (and design assumptions for each one) would need to be identified and documented. This would allow appropriate procedures to be developed to govern the appropriate behavioral constraints for each mode.

One unsafe control action has been chosen for the example: the initiation of an ITP climb maneuver when it is not safe to do so (context). The context states were identified by Thornberry as ITP criteria, ATC clearance, and clear airspace [190]. ITP-criteria satisfaction is communicated by the ITP-equipped flight computer. ATC clearance is a control action (CA-3) from ATC to the flight crew. Clear airspace is determined by the feedback to the flight crew from the weather radar and a visual scan for other aircraft. An abbreviated variable reference for the control structure is shown in Table 3-4 to complement the discussion. Note that only one type of control action (7a – initiate ITP) is detailed, although there are many possible control actions the flight crew may issue. CC-2 and IM-4 are included to highlight an important consideration. Even though those information links are not input directly to the flight crew, they do reach the ADS-B. Because the ADS-B does not itself issue commands, any communication it receives must be considered when analyzing the entities to which the ADS-B provides feedback.

In STPA Step 2, causal scenarios that contribute to a UCA are identified. For the ITP example, several tables are presented next that summarize the causal scenarios identified using the original STPA report, Thornberry's method, and the new STPA-RC guidance described in Table 3-2. First, Tables 3-5 and 3-6 show the findings in the original report (2012) with Thornberry's (2014) findings [190], [238]. Thornberry presented his causal scenarios as revisions (not additions) to the original report. He simplified the findings in the information-availability section to relate directly to the three process model variables (i.e., ITP criteria, ATC clearance, and clear airspace). He introduced scenarios in the new observability section, which at the time included detection, interpretation, and attentional demand. For the process-model section, Thornberry again simplified it to be aligned with the process variables. He also introduced a scenario within the new action-affordance section. There, he emphasized the importance of the controller being made aware of an inappropriate affordance, a concept that has been integrated more explicitly in STPA-RC.

Although some scenarios from the 2012 report were absorbed into simpler statements for Thornberry's revision, they are worth noting in Table 3-5. In the information-availability section in light-blue text are references to data being displayed and/or monitored. These statements are in fact information-observability scenarios. In the process-model section in light-orange text are scenarios having to do with the currency of process states and completeness of feedback. Those are important scenarios that should be considered in the control algorithm section; it is important for the controller to know when to request updated information. All these scenarios are included in the STPA-RC causal scenarios presented next.

Tables 3-7 and 3-8 present causal scenarios that were added or reintroduced into the ITP analysis using STPA-RC (2016). The new scenarios supplement but do not replace Thornberry's 2014 scenarios. The unavailability of affordance feedback about ITP being inadvertently initiated is documented explicitly in the availability section. Other scenarios having to do with affordance feedback and conflicting feedback are also included for consideration. Scenarios are also added to information observability. Fleming's scenarios about information not being displayed or monitored are reintroduced, and additional scenarios concerning the attending of displays are added.

The process-model scenarios are divided by behavior, modes and motives, although this is optional. Information interpretation scenarios are also included here. While Thornberry included interpretation scenarios in his observability section, STPA-RC emphasizes that information pickup is part of the controller's observation loop and can be addressed in the process-model analysis. The mode scenarios cover various inappropriate mental models having to do with supervisory modes and component-operating modes and assumptions. The control-algorithm section adds to the existing scenarios in which the controller does not request or confirm required information; Fleming's scenarios regarding the currency of process states and completeness of feedback are reintroduced here. Additional information-request considerations are added, and an objective-priority problem is suggested.

Table 3-5. In-Trail Procedure Causal Scenarios, 2012

| ***UCA:*** | **Flight crew initiates ITP when process states 1-3 are not satisfied** |
|---|---|
| | **2012 STPA Report** |
| Availability | - Change in own velocity/altitude/bearing **not displayed**<br>- Change in other's velocity/altitude/bearing **not displayed**<br>- Proper aircraft identifier of others **not displayed**<br>- FC does not receive communication from ATC<br>- Flight crew does not receive traffic info from ADS-B<br>- Flight crew lacking information from ATC<br>- ATC approval not on communication **channel that FC is monitoring**<br>- ITP equipment provides criteria data too late<br>- ITP equipment gives incorrect or ambiguous state information |
| Observability | |
| Process Model | Flight Crew believes:<br>- that their climb/descent capability is greater than it is<br>- it **has all ADS-B data** for local traffic<br>- ADS-B data to be accurate when it is not<br>- ITP criteria to be different than it is<br>- communication protocols with ATC to be different than they are<br>- communication with nearby aircraft to be different than they are<br>- in a different understanding of individual crew member responsibilities<br>- weather/turbulence to be better than it is<br>- ITP request to be approved when it is not<br>- ATC **approval to be recent** when it is old |
| Algorithm | Flight Crew:<br>- does not correctly check airspace is appropriate<br>- does not correctly check that all ITP criteria are met<br>- begins executing ITP prior to receiving approval<br>- delays in executing ITP after receiving approval<br>- does not re-verify conditions not changed after approval |
| Action | |

Table 3-6. In-Trail Procedure Causal Scenarios, 2014

| ***UCA:*** | **Flight crew initiates ITP when process states 1-3 are not satisfied** |
|---|---|
| | **2014 Revisions** |
| Availability | Any of the ITP criteria (PS 1.1-1.10):<br>- are incorrect, missing, not provided in the appropriate amount of time<br>- are in conflict which leads to an ambiguous ITP criteria<br>ATC clearance (PS 2):<br>- is incorrect or missing, not provided in the appropriate amount of time<br>- no longer remains valid (i.e. not refreshed in the appropriate amount of time)<br>Any clear-airspace indication:<br>- is incorrect or missing, not provided in the appropriate amount of time<br>- is in conflict which leads to an ambiguous airspace model<br>Conflict between PS 1, PS 2, and PS 3 |
| Observability | Any of ITP criteria OR their changes/updates:<br>- are not detected or not interpreted correctly<br>- take too long to detect and interpret correctly<br>- require too much attentional demand to detect and interpret correctly<br>ATC clearance or any change or update:<br>- Anything but ATC clearance is detected and interpreted as a clearance<br>- A revocation of ATC clearance is not detected and interpreted correctly<br>Any clear-airspace variable:<br>- is not detected or not interpreted correctly<br>- takes too long to detect and interpret correctly<br>- requires too much attentional demand to detect and interpret correctly<br>Conflict between the flight crew's interpretation of the three process variables |
| Process Model | Flight Crew believes:<br>- ITP Criteria has been met when it has not<br>- ATC clearance to be valid when it is not<br>- Clear-Airspace model to be satisfied when it is not |
| Algorithm | *No revisions* |
| Action | Flight Crew inadvertently affords the initiation of ITP and isn't made aware of this through feedback |

Extrinsic factors are all new, and example scenarios are presented in light of the ITP concept. Attentional demand is again emphasized, and ergonomics and human variability issues are included. Influences are notional for this example, but the scenarios included there demonstrate the usefulness of examining socio-organizational factors. As is evident, not all scenarios or suggestions can be very detailed at this point. The specificity of scenarios is contingent on how well the design and operating-philosophy of the system have been defined. As the concept evolves into a detailed design, the STPA-RC results can be updated as part of the systems-engineering process.

This example highlights some of the additional causal scenarios that might contribute to hazards that are possible using STPA-RC. The guidance developed here allows analysts—even if not experienced in human-engineering—to consider many aspects of human controllers with much more refinement than previous methods, based on the specific design and use of the system. However, there are admittedly even more potentially missed scenarios in this example that can only be realized by engineering and operations SMEs familiar with the design, philosophy, and socio-organizational aspects of the ITP concept.

Hazard analyses using this newly developed extension are inclusive of all the previous controller models and analyses while adding granularity to the Step-2 analysis. The gaps in STPA have been addressed, namely expanding the scope of the process model, adding fundamental human-engineering considerations, and including influences such as policy. All of these additional considerations remain grounded in STPA as a top-down approach. That means that the methodology retains a system-theoretic viewpoint. Safety constraints are identified at the system level, hazardous behaviors that violate those constraints are found, and then specific guidance is applied in the context of explaining those unsafe behaviors.

STPA-RC can be applied to any stage of product development. The next chapter focuses on the test stage of a product or system. The formats and analyses discussed so far are carried through where appropriate, as nothing about STPA is fundamentally different when the extension is applied. Because humans and software are becoming more highly coupled in modern systems, testing will become more and more difficult; using a hazard-analysis method that addresses the human controller with the refinements needed to explain intelligent behavior will be a benefit to any analyses and planning activities that support safety during test.

Table 3-7. In-Trail Procedure Intrinsic Causal Scenarios, 2016

| ***UCA:*** | **Flight crew initiates ITP when process states 1-3 are not satisfied** |
|---|---|
| | **2016 Additions (Intrinsic)** |
| Availability | - No feedback available that ITP was inadvertently commanded<br>- No feedback available that ITP was not disengaged when it was meant to be (affordance feedback assumed to be enough)<br>- Visual weather (out the wind screen) does not match weather radar (or digital weather report) |
| Observability | - Change in own velocity/altitude/bearing not displayed<br>- Change in other's velocity/altitude/bearing not displayed<br>- Proper aircraft identifier of others not displayed<br>- ATC approval not on communication channel that FC is monitoring<br><br>- Feedback (e.g., confirm tone) that ITP was commanded not heard by flight crew<br>- Pilot cannot see other pilot arm/disarm ITP or autopilot<br>- Wrong display read to determine distance to reference aircraft |
| Process Model | *Behavior*<br>- FC does not recognize bad weather on their attended displays (info pickup)<br><br>*Mode – FC believes:*<br>- ITP criteria or ATC clearance met due to poor interpretation (info pickup)<br>- that their climb/descent capability is greater than it is<br>- communication protocols with ATC to be different than they are<br>- communication with nearby aircraft to be different than they are<br>- in a different understanding of individual crew member responsibilities<br>- Autopilot is disengaged and will not climb to bug-altitude if ITP criteria not met<br>- ADS-B is in constant refresh mode when it requires manual interrogation<br><br>*Motive*: e.g., shortening transoceanic flight times is a high airline priority |
| Algorithm | *Controller does not request or confirm required data from process*<br>- Flight crew believes it has all ADS-B data for local traffic<br>- Flight crew believes ATC approval to be recent when it is old<br>- Flight crew does not correctly check airspace is appropriate<br>- Flight crew does not correctly check that all ITP criteria are met<br>- Flight crew does not re-verify conditions not changed after approval<br><br>*Controller prioritizes wrong objective:* Efficiency vs verification of safety |
| Action | *No additions* |

Table 3-8. In-Trail Procedure Extrinsic Causal Scenarios, 2016

| **Flight crew initiates ITP when process states 1-3 are not satisfied** | |
|---|---|
| | **2016 Additions (Extrinsic)** |
| Workspace | *Workload*<br>- Monitoring any process variable requires too much attentional demand to detect and interpret correctly<br><br>*Ergonomics*<br>- ITP disarm lever difficult to reach |
| Variability | - Pilot unable to read or hear certain types of displays<br>- Fatigue/crew rest, injury, etc. |
| Influence | *Behavioral Standards*<br>- Flight crews that achieve better mission efficiency seem to be getting assigned more preferable shifts<br><br>*Rules and Techniques*<br>- Policy memo about ADS-B refresh problems did not reach all pilots<br>- Flight crews developing habit to arm ITP before confirming criteria met<br><br>*Settings and Configurations*<br>- FMS has outdated set of desired mission altitudes from dispatch |

# Chapter 4

# Systems View of Testing

> "Plans are worthless, but planning is everything."[1]
>
> —*Dwight Eisenhower*

The second research objective is to *provide a common framework for test-safety planning that addresses both the safety of the test process and inherent system safety*. The following are the existing gaps in developmental product testing prior to this research:

- There is inconsistent expert knowledge at any given test-safety review board.
- There may be minimal expertise in new technologies (e.g., software, autonomy).
- The test-safety planning process does not use common visual aids in its documentation.
- Test engineers do not have a consistent method of tracing undesirable behavior or potential design flaws to effects on the system within the context of field use; this especially affects human-engineering experts, who cannot ignore the relationship between operating philosophy and system design.
- Problem reports tend to be reductionist (e.g., manufacturing error, component defects) and do not consistently explain system impacts through anything but written narratives.
- System-Theoretic Process Analysis (STPA) control structures do not acknowledge an explicit product testing stage to capture the particular sociotechnical dynamics found in a test enterprise.

From 2013 to 2015, I made several visits to the Air Force (AF) developmental-test (DT) enterprise at Edwards Air Force Base (AFB) in support of socio-organizational research conducted by Draper Laboratory. As an acquisitions officer in the AF, I was familiar with the profession and personnel of the AF test community upon joining this research. My background and contacts allowed for the Draper team to interact with offices, personnel, and equipment, including opportunities to observe real-time unmanned vehicle (UV)

[1] Remarks at the National Defense Executive Reserve Conference, 14 November, 1957.

missions. During this period, the team used contextual-inquiry methods to understand test practitioners' work activities and challenges through fact-finding interviews and active observations, and I was able to leverage some of those outputs in support of data for this thesis [44].[2] I either conducted or was involved directly in 40 interviews and discussions. Two were conducted over the phone with product-acquisition managers elsewhere in the AF, and the rest were on-site at Edwards: three were with people in test leadership, five were in the base safety office, six were airspace and airfield managers, two were responsible for range and mission-support activities, four were senior discipline engineers, and eighteen were front-line test planners, mission controllers, and aircrew.

I used knowledge gained from this work to perform a thorough organizational and policy review and construct an Explicit-Influence Map (EIM) for AF DT practitioners, as well as a detailed organization chart, which is presented in an abbreviated format in the next section. After gaining an understanding for how various safety policies affect product testing and how testing fits within a system lifecycle, I updated Leveson's generalized example of an organizational control structure (Figure 2-7 in Chapter 2) to account for the fundamentals of a formal product-testing stage with new inclusion criteria. I investigated a real-world flight-test project using STPA. This involved developing a test-safety planning format for STPA; I compared the STPA-based test-safety plan for the project to the traditional test-safety planning document produced by Edwards practitioners. This was done both objectively as well as through a human-research study in which survey participants were asked to indicate their preferences for either of the methods over several questions assessing intelligibility, informativeness, and implementability. Results and discussions for these tasks are presented throughout the remainder of this chapter.

## 4.1 Modern Test and Evaluation

AF Materiel Command (AFMC) oversees all three stages of an AF system's lifecycle (research/design, DT, and field-use/sustainment). As discussed in Chapter 2 (Figure 2-12), each product-lifecycle stage may also be divided into segments. The fielding stage of a system might begin with an initial evaluation—putting the system into real-world conditions by experienced users—and then progress to full-time field use. In contrast, the DT stage is more methodical. First, thorough *test planning* must occur to determine the aspects of the system that must be assessed [220], [239]. It is during this segment that both a technical and safety strategy are documented for the ensuing *test conduct*.[3] Finally, *test reporting* must occur in a manner that communicates DT findings for the fielding stage (or problems to the designers) [241], [242]. The processes of each of these segments are

[2] The focus of the Draper project was to study the RQ-4 Globalhawk (UV) test community to model operator use philosophies, methods of interpreting displays to comprehend system and problems, team dynamics, decision-making strategies, and operators' trust of the software. I was focused more broadly on test-safety planning, the organization's balance between test and test support activities, and the regulatory requirements for safety across the various disciplines in the organization.

[3] It is also during the planning segment that modifications are made to existing systems to incorporate the new or upgraded items being evaluated and the test-specific devices and instrumentation needed for the evaluation [240].

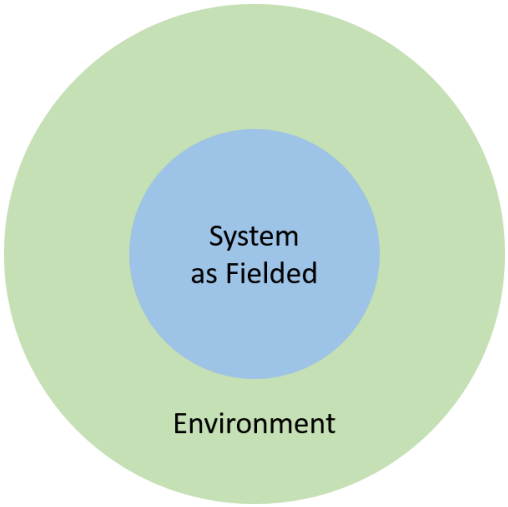



Figure 4-1. Representation of System in Fielding Stage

unique. They may be all performed by the same practitioners or each by a different set of personnel, depending on how the test enterprise is organized. In the AF, the same personnel are typically involved with all the segments of the test stage.

During any development stage, a product should be viewed as a system with a boundary to an environment. The system includes not only the physical product, but the entire sociotechnical organization that supports its operation. A non-structured system-theoretic representation of a fielded system might look like Figure 4-1. The system, shaded in light-blue in the figure, includes all aspects of the design, personnel, use-philosophy, and organizational structure that enables its mission. Because it is an open system, the environment (which is shaded green in the figure) shares with it a permeable boundary through which process inputs, outputs, and disturbances are conveyed.

In the test stage, the traditional view has been that the product being evaluated is isolated from the real environment by creating an artificial *test environment* [23], [243]. This view is represented in Figure 4-2 (a). The new product (or upgrade) is called the system under test (SUT), and the test environment is adjusted to isolate the SUT from the variability and uncertainty that the real world presents.[4] Although the ability to sanitize test conditions is one of the fundamental principles that differentiates DT from initial field evaluation, only considering the SUT is a reductionist viewpoint. It views the system as only the physical product and does not include personnel, test and support equipment, or the organizational management structure that enables the test mission. Operators are seen as supervisors of the system, a view that evokes the user-centered perspective discussed in Chapter 2. However, as indicated by the visible gap in the test environment's representation in the lower-right portion of Figure 4-2 (a), the real environment can never be completely

[4] If the efforts of DT and field evaluations need to be combined, the test conditions may be tuned to allow more of the real environment through or to emulate more aspects of the real environment. In combined evaluations field users may be invited to assist in DT operations.

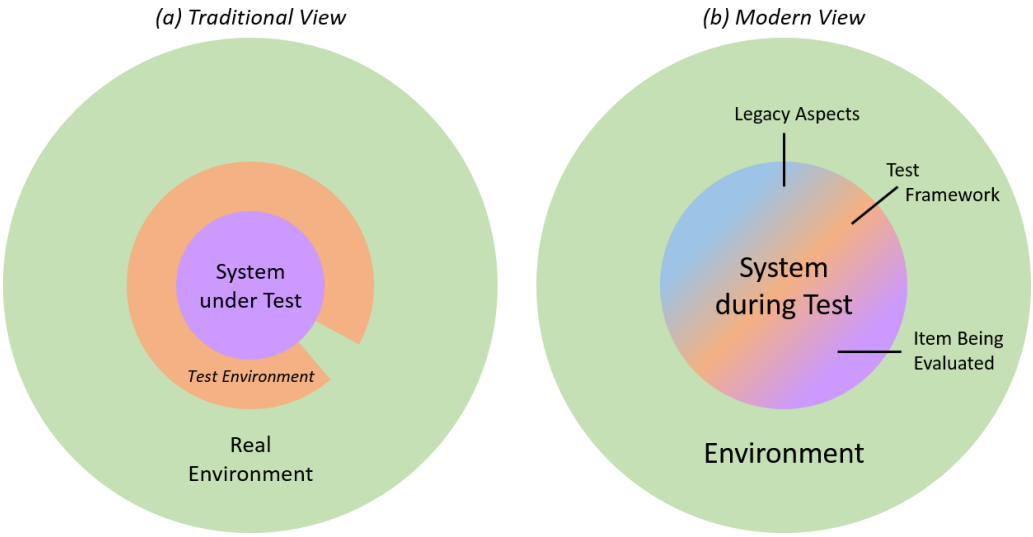



Figure 4-2. Representations of System in Test Stage

prevented from interacting with the product, no matter how isolated the test enterprise attempts to make it.

A more modern view for the test stage is proposed as shown in Figure 4-2 (b). This view better resembles Figure 4-1. There is only the system and the environment. A more appropriate term in this view is *system during test* (SDT). Here, the system in consideration once again represents the physical product, personnel, use philosophy, and organizational structure. Anything under the control of designers, practitioners, and organizational stakeholders is part of the system (including airspace/ranges, support assets, and policies). The SDT is shaded in three colors in the figure. Light-blue represents aspects of the system that are identical to—or emulate—aspects of legacy fielded systems. Purple represents aspects that are new or upgraded. These can be components, software, tactics, procedures, and even control modes that are under evaluation. A more appropriate term than SUT for this purple region is *item(s) being evaluated* (IBE).[5] Orange represents aspects of the system that exist to support the testing. These include software, apparatuses, targets, instrumentation, policies, and techniques that provide technical data to support the goals of the evaluation and/or control for safety.[6] A term for these aspects that is more appropriate than test environment is *test framework*.[7]

The SDT encompasses all three of these aspects. A non-structured representation such as Figure 4-2 (b), however, does not describe the specific entities within the system and identify what is considered legacy, IBE, or test framework; a better way to represent those details would be to use a functional control structure, which is demonstrated in

[5] The word *item* in IBE is intended to be generic and does not imply that what is being evaluated can only be a physical component. If a product is completely new and not just an upgrade, the IBE can encompass a large portion of the system with little to no legacy aspects.
[6] Goals include evaluating performance, verifying models, or confirming design assumptions.
[7] "Environment" implies that something is not controllable. However, the test framework is within the control of the system, even if part of its purpose might be to emulate aspects of the real-world environment.

Section 4.3.3. The important thing to recognize here is that regardless of what lifecycle stage a system is in, it includes all the aspects of the sociotechnical enterprise. There is no isolated SUT shrouded by a test-support infrastructure; instead, everything including the test framework *is* the system. All the entities within it may have some interface with the environment, via intentional processes and/or disturbances.

Summarized from Chapter 2, the two safety goals during the test stage are:

A) Determine safety of the system as designed and intended for use
   - Confirmation that design-stage models (e.g., computational dynamics) used adequate assumptions and input parameters
   - Risk reduction for aspects of the system with no accepted models (e.g., human or autonomous components)

B) Ensure safety of the testing process itself
   - Test techniques, configurations, instrumentation, range support
   - Buildup approach when verifying models that might be inaccurate

The first goal, safety of the inherent product, can be seen in Figure 4-2 (b) as ensuring that the legacy (light-blue) and IBE (purple) portions of the system have been configured to accurately represent the manner in which the system will (or might) be fielded. When components are still immature or for some reason unable to be configured completely, those limitations must be documented. The behavior and performance assumptions for new components or capabilities must also be documented. Ideally, the design stage provides these assumptions in the form of models, model outputs, and bench-testing results. Predictions that could not be made during the design stage—particularly regarding human and software behavior within the greater system—should also be documented because the test stage will be where risk reduction must be planned to account for those knowledge gaps. The design stage would ideally also perform a system-theoretic hazard analysis, like STPA, to identify unsafe control actions (UCA) and causal scenarios for the system as it is intended to be fielded. The assumptions in that analysis about the safety control should be analyzed during the test stage.

The second goal is the safety of the test conduct itself, which affects testers, support personnel, and the public near test facilities and ranges. This consideration is often called *test safety*. The test-framework (orange) aspects of Figure 4-2 (b) come more into the focus, and the complete SDT is analyzed for test-specific characteristics that contribute to hazards. Often the test techniques, instrumentation, and support assets introduce new scenarios for hazards in the system, and the designers' hazard analysis need to be updated by test practitioners when they receive the system. Additionally, the risk-reduction efforts from the first safety goal are inherently hazardous in light of the second goal, so what is

commonly referred to as a *buildup approach* is used as appropriate when test planning [220].[8]

It is common for the two factors that justify test safety (unique evaluation techniques and risk-reduction buildup) to apply to the same test project. An example would be a new UV going airborne for the first time. The sorties performed to accomplish the initial flight envelope expansion follow rigorous test-safety procedures until the vehicle structure, propulsion, and flying qualities are proven. This maturity determination—verified consistency and integrity of the performance of the basic vehicle outputs—usually allows for the basic aircraft and flight controls to be subsequently managed by looser safety restrictions. Following that, the rest of the system's capabilities may be evaluated, governed by test-safety principles that are appropriate for the techniques used.

During the test stage, planning for test safety often has priority over reporting on aspects of inherent product safety, because in the short term the health and equipment of the testers are what is at risk. In AF DT locations, test safety is governed by its own specific policies [24], [244]. The test-safety planning function not only includes a hazard analysis on the testing of each new system, but also establishes general best practices for how tests are planned and executed. Safety in the field is more concerned with *flight*, *ground*, and *weapons* safety, operating practices that govern procedures and planning for all general activities. Test practitioners, while evaluating unproven systems in the test stage, must employ these operations-safety paradigms in addition to test safety.

As systems become more complex and include more human and software interdependencies, it will become increasingly difficult to apply the older view in Figure 4-2 (a). Particularly with UV systems, findings during testing will update knowledge and procedures for flight safety every bit as much as test safety. In the above UV risk-reduction example, the performance of the flight computer, even before the basic vehicle performance is demonstrated, is already crucial to the flight safety of the system. Sometimes many immature technologies and/or novel software architectures must be tested concurrently. It can be difficult to characterize or test the performance of a single feature on a UV that is already considered safe to operate under normal flight rules.

The current AF test-safety planning method relies on senior engineers and operators around a table drawing from their experience to identify the most plausible chains of events that could lead to accidents. This technique—akin to storytelling—is sometimes supplemented by modeling and simulation, probabilistic calculations, or other data from the designer. Often the final safety plan document is built by copying the plans from previous similar tests, making updates, estimating likelihoods of hazards, getting approval by a committee of experts, and presenting it to leadership in a probabilistic format for risk acceptance. When new or exotic technology comes along (e.g., autonomy), the limitations of the traditional method begin to become apparent.

[8] Typically, brand new flight systems begin with aerodynamic and structural testing, gradually increasing airspeeds, altitudes, vehicle configurations, maneuvers, and flight durations over subsequent sorties. Then avionics, radios, and other onboard capabilities are evaluated. The aggressiveness of the buildup is often a tradeoff made with schedule and cost constraints and negotiated with the AFMC program office.

As a product is developed throughout its lifecycle, practitioners at the test stage should reference not only their own operating experiences, but more importantly the control model that was created by designers for the specific system as well as their initial hazard analysis. Testers can update this model to reflect the SDT, including adding the test-infrastructure and the test stage's socio-organizational influences. Hazard analyses can also be updated to account for test activities. Throughout the test planning or test conduct, discrepancies that are found in the designers' intent can be communicated in a common language if practitioners at all stages have adopted a common modeling format.

System-Theoretic Accident Model and Processes (STAMP) is proposed as that common modeling format. The next section discusses the AF test enterprise and use the AFMC lifecycle stages as a template to update the generalized example of the STAMP organizational control structure. The new aspects of the updated STAMP diagram are discussed as they pertain to the concepts introduced above. Using STPA to perform hazard analyses in the test stage addresses the first three research gaps in this chapter. It provides a visual aid for planning (hierarchical control structure) and standardizes the method in which unsafe behaviors are identified and traced so that the hindsight provided by experience is not the only aspect informing the analysis. The remaining gaps are addressed by updating the STAMP general control structure and adding guidance to the information that should be managed within and among a system's lifecycle stages.

## 4.2 The Organization

Understanding organizational and policy relationships is important for anyone attempting to create a safety-control structure, and a properly represented system includes socio-organizational entities [7]. The fact-finding interviews and policy reviews I performed to understand the AF test organization is a good beginning, but the people who can best use organizational knowledge to build a STAMP representation of a system are its regular practitioners. Any safety tools and products that evolve from my short-term understanding are intended to be updated and kept current by the stakeholders of the organization.

Section 4.2.1 gives a brief background of the test organization, while Section 4.2.2 discusses the different safety policies in the AF, including test safety. Section 4.2.3 examines these policies again but from the perspective of an EIM. An updated STAMP organizational control-structure example that includes a test stage is presented in Section 4.2.4, developed using the acquisition and test philosophies identified in this research.

### 4.2.1 Air Force Test Center

Edwards AFB is a unique site that had its government origins in the 1940s as a remote location for sensitive military flight research away from populous areas like Wright Field in Dayton, OH.[9] The work evolved over decades, transforming Edwards into a permanent military work location as a well as attracting commercial flight-research endeavors to the Mojave Desert. The AF has recently consolidated its formal DT enterprise, named it the

[9] Wright Field has since grown into Wright-Patterson AFB and houses AFMC.


Figure 4-3. Edwards Lakebed Runways (https://www.nasa.gov/)

AF Test Center (AFTC), and officially headquartered it at Edwards; however, until that announcement Edwards had always been the center of AF DT in everything but name [245]. AFTC leadership oversees thousands of test projects annually and employs tens of thousands of professional testers both locally and at various other locations around the globe. Work at Edwards, the home location, takes advantage of the dry lakebeds surrounding the main airfield runway stenciled with almost a dozen emergency runways.

Complementing Figure 2-11 in Chapter 2, Figure 4-4 shows the three main test wings that fall under AFTC. The Arnold Complex in Tennessee conducts wind-tunnel research, and the 96th Test Wing at Eglin AFB, Florida evaluates weapons and other battle effects that are employed by military air and space vehicles. While those two missions are important to AF DT, this thesis leverages the organization and practices of the 412$^{th}$ Test Wing (412TW). The 412TW evaluates air vehicles, their software, and their integration with weapons and enhancements.[10] It is co-located at Edwards AFB with AFTC.

I built an organization chart for the 412TW based on the information and documents obtained from local fact-finding interviews and inquiries. This product primarily focuses on physical and administrative relationships, including working groups, planning, and

[10] The design, integration, testing, and employment of weapons is not discussed in this thesis.

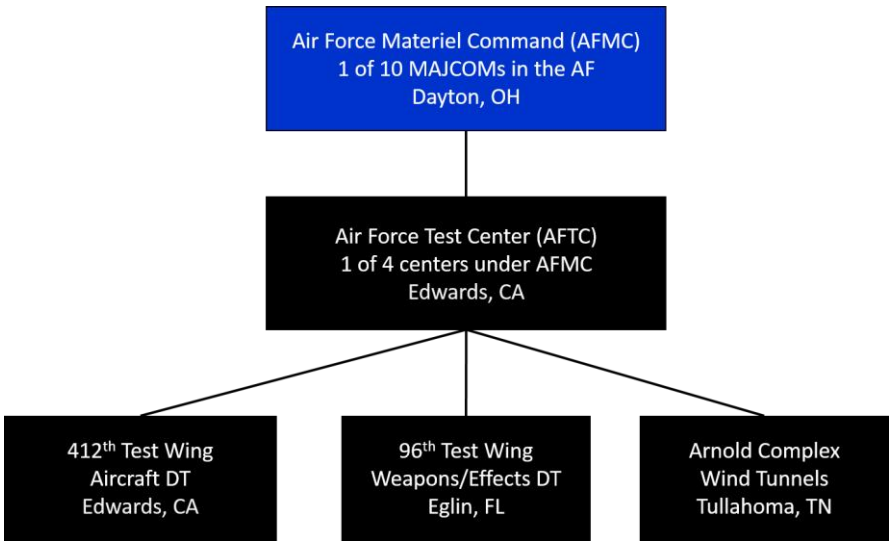



Figure 4-4. AFTC Organization

policy flows. Although the full organization chart cannot be reproduced in this thesis (as it contains official government information), an abbreviated version is shown in Figure 4-5. The organization chart can serve as a reference when building a safety-control structure. Other useful sources of information for the control structure are discussed in Chapter 2 [21].

The 412TW contains a set of safety offices indicated by the green box in Figure 4-5 that are connected to the Test Wing by a dotted line. Almost every level of command in the AF that is a wing or higher has a safety office.[11] A safety office determines how its corresponding command executes the policies and procedures required for the safety of its mission. While flight, ground, range, and weapons safety are managed at most AF locations, test and test-range safety are also managed at Edwards.[12] Additionally, a wing's local safety office serves as a liaison to the AF Safety Center (AFSEC) in New Mexico. AFSEC manages the best practices for the various safety paradigms in the AF, maintains incident-data repositories for AF products, and standardizes the training of safety personnel at all AF locations.

Below the test wing are various group-level offices. The *operations group* serves as the backbone of the test mission. It consists of several flight-test squadrons—typically one per airframe type[13]—as well as an auxiliary squadron that manages airfield and airspace activities. Typically, flight-test squadrons own, configure, and operate the IBE as well as legacy aspects of the product SDT. The airfield and airspace squadron performs activities that embody real-world air-traffic control (ATC) as much as feasible, but at

[11] Figure 4-4 does not visually shows it, but AFTC and AFMC have safety offices of their own.
[12] More discussion on the various safety policies in the AF follows in Section 4.2.2.2.
[13] The F-22 Raptor, F-16 Fighting Falcon, and F-35 Joint Strike Fighter each have their own flight-test squadron. All transport aircraft fall under one squadron, and all bombers fall under one as well. The RQ-4 Globalhawk squadron hosted the research team, and it has since then become the squadron responsible for managing all UV testing at Edwards.

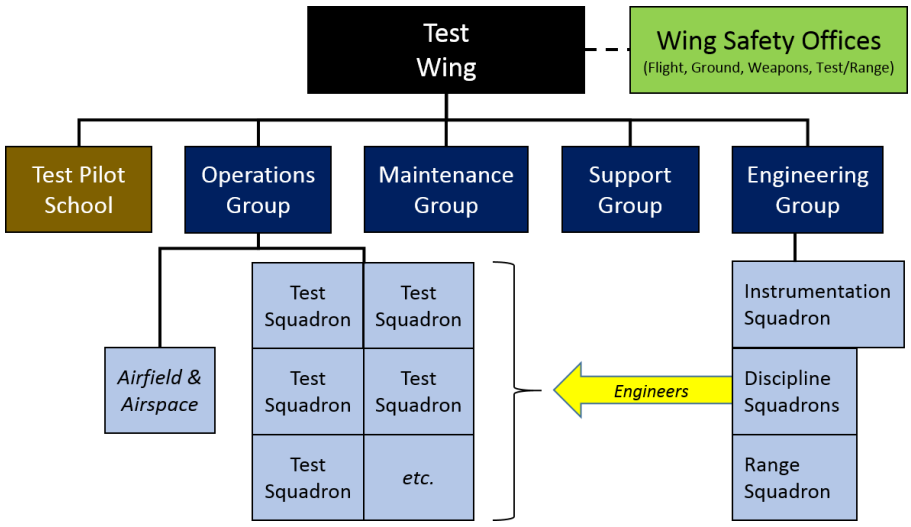



Figure 4-5. 412th Test Wing Organization

Edwards the airspace controllers have regulatory discretion to allow testers to utilize many flight techniques and airspace configurations not allowed in the field [246], [247].

The *engineering group* contains offices that manage the test framework and its integration with the IBE. The group includes various engineering-discipline branches containing expertise in everything from propulsion to flight controls to human factors. An instrumentation squadron works with the discipline experts to develop methods and tools to measure the required data for product evaluations. A range squadron owns and operates assets that support testing, such as ground targets, electromagnetic phenomena (including telemetry collection), and weapon spotting. Typically, members of the engineering group are placed in the flight-test squadron, either organizationally matrixed to them or as liaisons on an as-needed basis. The yellow arrow in Figure 4-5 represents this. While the flight-test squadron practitioners that report directly to the operations group are the ones who plan, train, and practice to operate the SDT, the engineering-group practitioners are truly the experts in what the IBE is and how best to measure its performance.[14]

The *maintenance group* works with the instrumentation experts to configure the system for testing and measurement [240]. Similar to maintenance phases in the field, maintainers in the test stage perform standard actions between operating phases. The *support group* encompasses many supplementary activities for the wing (such as information technology), and it manages the legal allocation of the electromagnetic spectrum[15] for daily wing operations. *Test Pilot School* (TPS) is not a group per se, but reports directly to the test wing. TPS trains test pilots and flight-test engineers with a year-long academic curriculum covering aircraft performance, flying qualities, avionics, and test management including training in aircraft and mission-control rooms. Graduates typically

[14] There are six verbs that may be used in planning to qualify how the IBE is assessed (observe, compare, demonstrate, determine, evaluate, and verify), and detailed definitions for each are available in test-planning references [220], [239].
[15] Radio and telemetry frequencies.

comprise the majority of the flight-test squadron practitioners that report directly to the operations group (including the squadron leadership).

The control structure for the test conduct for a specific project is discussed in Section 4.3. Before modeling safety-control structures, stakeholders should determine the phases that should be covered. phases are defined periods or units of work activity, and each is a socio-technical activity that may repeat or alternate with other phases (see previous chapter). For example, in the field-operations segment of the fielding stage, common phases include mission briefing, mission operations, mission debriefing, and system maintenance. The test-conduct segment of the test stage is the focus of this research, and here it is assumed to contain the phases of mission planning, briefing, test operations, debriefing, and maintenance. As similar as those are phases in the field, it is important to emphasize that every individual mission (or test event) covers a specific subset of maneuvers and evaluations. Unique parameters, settings, procedures, safety considerations, and objectives must be planned, briefed, and executed for every single test sortie.

Although the test-planning and test-reporting segments are not covered at length here, a good reference for the planning segment is Chung's work from 2012 to 2014 on lean engineering and test programming at the 412TW [248], [249].[16] As part of his research, he built a safety-control structure for test planning, reproduced in Figure 4-6. Although not exactly in the same visual format chosen for this thesis, Chung's control structure similarly labels control actions and feedback with numbers. An important thing to re-emphasize is that the control structure does not mirror the organization chart. Instead, the control structure functionally diagrams the active process.

Chung's model was able to address various phases within test planning, including test-plan authoring, plan editing, plan review, and test approval. The specific controls and feedback that facilitate those actions are left for the reader to reference [249]. Chung also performed STPA on the test planning segment and was able to determine non-optimal review sequences, identify resource pressure, warn against possibly conflicting multiple controllers, and determine areas where reviewers lacked qualification standards. Regarding the test-safety hazard analysis itself, Chung highlighted a lack of visual-planning tools as well as detailed safety-planning models that the design stage should be providing. Both of those findings are covered by research gaps in this thesis. Since Chung began his work, the 412TW has updated policies in order to consolidate how the different offices in the wing cooperate to plan ahead for new test projects, including required tasks for range, airspace, engineering, and operations personnel and a mandate to revisit the policies every two years [26].

Because of the proliferation of autonomy-capable military systems and concepts, AFTC is anticipating an increase in UVs flying at Edwards. The 412TW safety culture is

[16] Not all acronyms in the figure are important for this thesis, but "AFTC/SET" is the test-safety office, "OG/CC" is the operations group commander, "TW/CC" is the test wing commander, and "unit" refers to the flight-test squadron. In the test stage the safety office is intricately involved in the planning that precedes conduct, whereas in field-use, safety offices serve more as a background source of best practices and incident data and do not involve themselves in day-to-day operations.

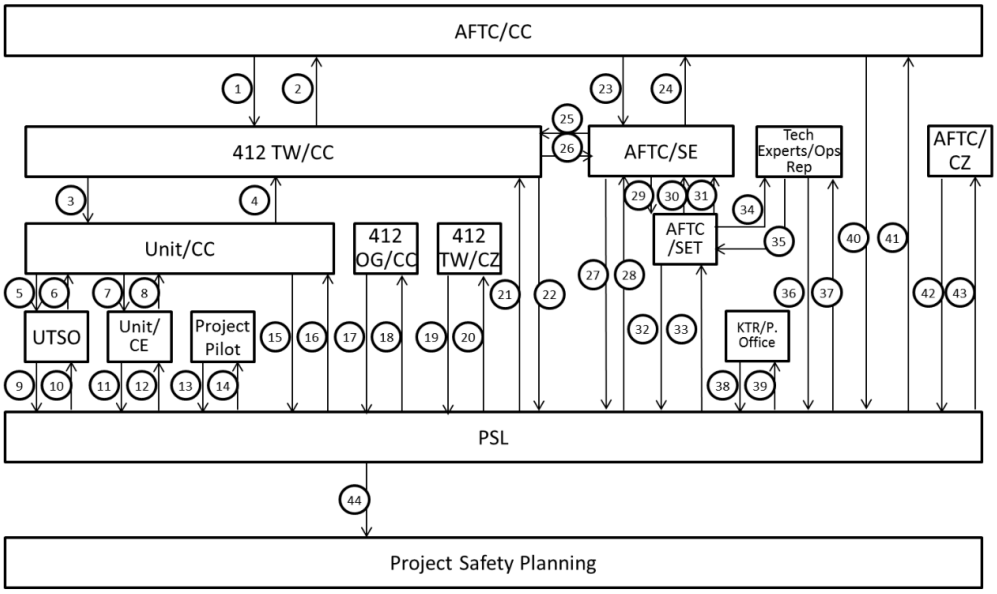



Figure 4-6. Planning-Segment Control Structure [249]

interested in modernizing the processes that affect airspace planning, range clearance, and flight-test procedures for these types of systems. Airspace managers, as a conservative precaution, are currently mandating complete spatial and temporal deconfliction between UVs and MVs by a relatively wide margin. However, available airspace is diminishing with the proliferation of products that need to be tested. Moreover, collaborative employment of both manned and unmanned systems will be necessary for future fielding concepts, necessitating an early understanding of planning for such circumstances during test.

The 412TW has already begun to reassess its criteria for determining UV system maturity [26]. The maturity determination governs how many additional test-safety redundancies must be in place as risk reduction is performed on a product. The legacy criteria for UVs was previously derived from that of MVs, basing the maturity determination on number of flights or flight hours. However, the practitioners now understand that the basic proven performance of the vehicle (evidenced by flight time) is only part of the equation. The sensing and deciding functions (such as what might be accomplished by the autonomous flight computer) will also be steadily proven to some level of capability and must be considered as well.

AFTC presents an opportunity for applying STPA to encourage the use of systems theory in the test-safety process. The DT culture is looking for a modern methodology to assess safety. New technologies present challenges.[17] Like many sociotechnical constructs that began in the last century and are coming of age with modern technology, flight-test

[17] The engineering group at Edwards does not yet have a distinct division for software engineering or autonomy. This may change soon with the emergence of not only UV technologies but cybersecurity initiatives. STAMP research at the Massachusetts Institute of Technology (MIT) has directly contributed to updated AF cybersecurity policies [250], [251].

practices are a combination of many decoupled and derivative regulations and procedures that have evolved naturally and are applied in parallel to ensure safe operations. People are relied upon to do their parts during planning and operations. The need exists to update current planning practices to model not only the new product, but also include the multi-dimensional and inter-organizational control structure at Edwards to capture the entire SDT.

### 4.2.2 Air Force Safety Management

Safety in the AF is managed through various standards, policies, and practices that all affect how a product is designed, tested, certified, fielded, and generally operated. This section covers the Department of Defense (DOD) standard for safety, MIL-STD-882, that affects almost every aspect of safety practice in the military, followed by an overview of those practices.

#### *4.2.2.1 MIL-STD-882*

MIL-STD-882, *System Safety*, guides the DOD systems-engineering approach for "eliminating hazards, where possible, and minimizing risks where those hazards cannot be eliminated" [11, p. 1]. It establishes a core system-safety process and describes the tasks that system developers should accomplish throughout the lifecycle of a product to implement safety during its development. In this way it serves as both a safety-engineering template and a safety-management guide.

The MIL-STD-882 *system-safety process* is as follows:

1) Identify the hazards
2) Assess the risk
3) Identify risk mitigation procedures
4) Reduce the risk
5) V&V the risk reduction
6) Accept any residual risk

That system-safety process has been in place since the beginning of the system-safety movement, and STAMP has its roots in its philosophy of identifying *hazards* and letting them guide the safety analysis and risk reduction efforts. In fact, the definitions of the terms *hazard* and *mishap* are very similar between STAMP and MIL-STD-882. However, the following excerpt from the latter reflects a distinction in how hazards are traditionally viewed in practice [11, p. 10]:

> Hazards are identified through a systematic analysis process that includes system hardware and software, system interfaces (to include human interfaces), and the intended use or application and operational environment. Consider and use mishap data; relevant environmental and occupational health data; user physical characteristics; user knowledge, skills, and abilities; and lessons learned from legacy and similar systems. The hazard identification process shall consider the

entire system lifecycle and potential impacts to personnel, infrastructure, defense systems, the public, and the environment.

While 882's edict for a *systematic* approach to its system-safety process has been implemented successfully in industry, the methods in which hazards are identified and risk is defined have historically been implemented by the chain-of-events view of accidents instead of a *systemic* view of accidents. In the former—reductionist—view, a hazard is seen as a predictable condition that can be mapped from the bottom-up with event trees, fishbone diagrams, or other forms of root-cause analysis. Probability of the ensuing mishap is then either calculated using a formal Probabilistic Risk Assessment (PRA) method like a Fault Tree Analysis (FTA), or it is estimated via informal methods such as in a Preliminary Hazard Analysis (PHA) [143]. Traditionally the risk level is then seen as a function of the probability and consequence (or severity) of the mishap.

The ubiquitous "risk matrix" familiar to most engineers and system managers is reproduced from MIL-STD-882 in Table 4-1 [11, p. 12]. The mishap severity categories are defined using specific criteria covering life, injury, and cost, while the probability categories are defined using qualitative or quantitative thresholds of likelihood. While MIL-STD-882 has recommended values for these criteria and thresholds, most organizations establish their own values within their safety policies. The EIM for the 412TW (Appendix B) contains 41 documents that include or reference the risk-matrix concept, some with their own established criteria and thresholds [11], [23], [24], [27]–[30], [35], [203], [224], [234], [243], [244], [252]–[279].

The MIL-STD-882 *management tasks* (which in the AF are overseen by the AFMC program offices) are mandated to be performed by system managers as of the latest version (E) of MIL-STD-882. The tasks are shown here (subdivided into lifecycle and associated tasks) with the system-safety processes that each task supports, where applicable:

- Lifecycle tasks
  - Preliminary Hazard Analysis (PHA) : Establish hazard documentation early in development (system-safety processes 1–3)
  - Requirements Analysis: Fix requirements or mission use concept to eliminate or reduce hazards (system-safety processes 3–4)
  - Functional Hazard Analysis (FHA): Map known physical components to functions to establish probabilities of failure (system-safety processes 1–3)
  - Subsystem Analysis: Allocate safety requirements based on FHA through detailed design (system-safety processes 1–4)

Table 4-1. Traditional Risk Matrix [11]

| SEVERITY / PROBABILITY | Catastrophic (1) | Critical (2) | Marginal (3) | Negligible (4) |
|---|---|---|---|---|
| Frequent (A) | High | High | Serious | Medium |
| Probable (B) | High | High | Serious | Medium |
| Occasional (C) | High | Serious | Medium | Low |
| Remote (D) | Serious | Medium | Medium | Low |
| Improbable (E) | Medium | Medium | Medium | Low |
| Eliminated (F) | Eliminated | | | |

  - System Analysis: Verify compliance with safety requirements and identify new hazards associated with integrated design (system-safety processes 5–6)
  - System of Systems Analysis: Ensure system does not introduce risk when interfacing with other systems (system-safety processes 1–6)

- Associated Tasks
  - Operations and Support: Examine the system-use workflow and procedures (how it is operated)
  - Health Hazard Analysis: Ensure system complies with occupational safety and health (OSH) regulations
  - Test and Evaluation: V&V system risk reductions and manage local risk for test events using system-safety process
  - Mishaps and Problem Reports: Review data from product or similar products to update the hazard analysis

Although the system-safety process has remained a fixture since the early 1960s, the management tasks only came to be ten years later when the first iteration on 882 was codified. Subsequent revisions (A–E) have seen the tasks added, modified, deleted, and restored throughout the last five decades.[18] Evident in the language and structure of

[18] MIL-S-23069 came first in 1961 due to the growth of ballistic missile technology and the system-safety movement, and fault tress were the primary method of identifying hazards; MIL-STD-882 was published in 1969, made mandatory for the entire defense department, and codified the management tasks; revision A in 1977 defined risk as a function of probability and severity of a mishap (risk matrix) and mandated that risk be calculated and accepted by system managers; revision B in 1987 added software considerations and increased management tasks; revision C in 1996 combined hardware and software tasks; revision D in 2000 removed all management tasks due to acquisition reform; however, revision E in 2012 restored management tasks [48].

*lifecycle* management tasks is the strong influence over history from chain-of-events-based safety-certification techniques that evolved alongside the standards. Leveson et al. report that Aerospace Recommended Practice 4761, a safety assessment process that supplements civilian airworthiness certification of aircraft, uses FHA, a technique that includes fault trees and subsystem analyses [184]. Military airworthiness certification processes, discussed more in Section 4.2.2.3, also prescribe techniques like Failure Modes and Effects Analyses (FMEA) that support FHA and other 882 lifecycle management-tasks [28].

Properly using STPA to support system development meets the intent of the MIL-STD-882 lifecycle management-tasks. A system-theoretic hazard analysis views a system in its entirety from the beginning of development. Requirements and safety constraints would always be revisited throughout the lifecycle, and the analysis would be updated by testers and then by field users.

The *associated* management-tasks are met by a systems approach as well. Operations and support is analogous to the concept of phases and how the regular, repeating activities at the unit-of-work level can affect each other over multiple phases. OSH and workplace safety are important and should be managed as well. Safety during the test stage is also mandated here. Its objective, mentioned several times in this thesis, is to "eliminate [or reduce] the hazards…for both the system and the test events [e.g., test safety]" [11, p. 82]. Finally, mishap investigations and problem reports are required for both incorporation of past lessons into the system design (and use) and indicators of potential hazardous behavior.

For AF testers, system safety means more than just the reduction of risk performed on the one or two test vehicles they operate for the test project. With the modern hazard-analysis approach offered by STPA, the potential exists for DT to contribute more to product-safety determinations. This goal requires the use of STAMP concepts applied to a formalized test stage.

#### 4.2.2.2 Traditional Safety Practices

Safety throughout the development, employment, and sustainment of AF systems is governed by many disparate practices and information databases (see Section 4.2.3 for how they visually map to the test enterprise). Some of these practices, such as problem reporting, directly support a product's lifecycle management. The rest are more general, covering some but not all aspects of a complete system. The MIL-STD-882 system-safety process (particularly the risk matrix) is referenced in several places, as noted in the previous section. The tools and procedures that have evolved to fulfil these safety practices are often rooted in legacy, and authority to update them rests above the level of the test enterprise.

Some specific safety practices are covered in the next few subsections, including sample references. In general, there are guidelines for occupational safety, risk management, and mishap prevention that exist as high-level AF policies. These guidelines affect all the specific practices as well as all planning and operating phases in the AF [29], [30], [254]. They also establish the responsibilities of AFSEC and shape everything from workplace safety to motorcycle-riding policies (subset of ground safety) to hazardous-

material handling rules. Mishap-prevention guidance in particular has some very direct influences on test safety, which is discussed in its own subsection below.

*Operations Safety*

Safety during the conduct of an operating process, whether it be during DT or fielding, is covered by several techniques and best practices which generally fall under an umbrella referred to here as operations safety. These are general risk-management initiatives to reduce the likelihood of hazards, and they include flight, ground, weapons, and—in applicable locations—range safety rules.

Airspace rules exist to establish the guidelines for procedural and positive control of aircraft by ATC, and include Federal Aviation Administration regulations as well as AF and military operating area procedures [32], [33], [246], [247], [280], [281]. AF-specific training guidelines and general flight rules and procedures also exist to establish common standards for the methods that aircraft use to conduct their activities when sharing a multi-dimensional air and battle space [31], [282]–[285].

At Edwards, for multiple vehicles to share the same airspace requires a capability called "see and avoid"; this is a legacy aviation concept that means all flight vehicles must be capable of avoiding conflicting traffic without a prompt from ATC, based on timely visual detection of conflicts [25, p. 113]. No legacy AF UV is considered capable of achieving this, and newer systems that might be capable will need to be evaluated with a strategy that can prove the maturity of both the detection and avoidance technologies.[19] The 412TW maturities discussion mentioned previously must consider this strategy, but at the same time practitioners must question if the legacy airspace rules are appropriate for newer systems. Emerging technologies might benefit from more efficient methods and control strategies for ensuring that two vehicles do not get too close to each other.[20]

Another legacy underpinning that might be overly constraining UV systems is ATC radio protocol. Even when a traffic command is given to a UV human operator—who supervises the vehicle from a remote ground control station (GCS)—the syntax in which the advisories are given must assume the human is in the UV (e.g., "you have traffic 10 o'clock, high, 4 miles"[21]). If the UV were indeed flying in shared airspace, the human controller would have to undertake additional mental effort to determine how to search for and avoid the other traffic, because his visual perspective is not the same as the egocentric perspective of a pilot inside an aircraft. However the radio protocol remains the same regardless of whether the vehicle is manned or unmanned. Ironically, the ATC voice radio signal is often received physically by the UV, then encoded and transmitted to the GCS, so that the human operator can then send guidance commands back to the UV. Future work in providing more efficient methods of managing traffic flow and giving advisories is also

[19] And the decision-making, to be comprehensive.

[20] STAMP research at MIT is investigating solutions for this problem [286].

[21] Advisories are always given in the order of azimuth, elevation, and distance from the perspective of the cockpit pilot (if the pilot were to exist) facing forward in the aircraft. "10 o'clock" would be an azimuth 60 degrees left of forward.

needed. A STAMP safety-control structure can easily foster hazard analyses for these types of challenges.

Military-range safety and operating procedures exist in the form of guidance by the joint Range Commander's Council (RCC) and, pursuant to that guidance, policies at the DOD and AF levels [27], [34], [198], [224], [264]–[266]. Range safety is deeply invested in PRA, reaching back to legacy range safety traditions from the space age as discussed in Chapter 2. There is no commonality among the many military ranges in the government for the tools and methods that are used to model and calculate the probabilistic casualty estimates which are required to approve airborne items on ranges.[22]

Operating standards also exist specifically for every type (or class) of AF mission system [287], [288]. Each has its own methods as well as technical manuals that aircrew must study, train with, and sometimes use as a reference during operations. Use of manuals includes memorizing boldface procedures for urgent or critical system conditions. Furthermore, tactics, techniques, and procedures (TTPs) exist for each system and the types of missions it is capable of performing. These policies derive from military doctrine and are typically confidential documents.

Crew resource management (CRM) training is presented regularly to aircrew members to encourage the appropriate use of communications during operations and the recognition of mentally deficient states in oneself and one's teammates [36]. The concept was developed in the 1970s in an effort to reduce human error, which at the time was already receiving the vast majority of the blame for aircraft accidents [233], [289]. Another practice, called operational risk management (ORM), occurs right before a sortie; all crew members must fill out a questionnaire regarding several factors that studies have shown to be correlated to higher accident rates (e.g., how much sleep did the member get the night before) [30]. If the mission scores too high of a quantified risk level based on the operators' answers, local supervision must be consulted for execution approval.

*Accident Investigations*

Some incident investigations are reviewed for discussion in Appendix A. There is a legal requirement for an independent accident investigation board (AIB) to examine a mishap if it meets certain injury and cost thresholds [290], [291]. The Air Force also has a separate safety investigation board (SIB) process that has additional access to information that is legally privileged to only the flying community [35]. Those reports contain findings made in the light of practitioner culture, influences, and procedures, maintaining more user semantics. Even then, there may still be causal scenarios for hazards that are tucked in the discussions mid-way through the reports or perhaps only elucidated through in-depth inquiries that occur after the official investigations. Without a system-theoretic model of

[22] Most of the tools evolved from the same mindset as the regulations; namely, they are meant for ballistic objects with limited propulsion, gliding, and guidance capabilities. During my research, I learned of at least six methods and tools used singly or in combination for calculations. One of those tools, "Weapons Danger Zone", is mandated in the main AF range regulation [34]. It is not easily capable of being modified to assess new items or configurations that are not already fielded and modeled.

the system and a pre-established hazard analysis, the lessons found by exploring an incident are difficult to put into context, resulting in quick-fix solutions based solely on hindsight bias [146], [230].

An evolution of human-error modeling called the Human Factors Analysis and Classification System (HFACS) became a requirement for accident investigations in 2009.[23] HFACS is used to label types of errors, problems, or inappropriate decisions made by humans and organizations in order to produce historical data on error classes and improve preventative training [292]. The analysis is based on Reason's "Swiss Cheese", a chain-of-events accident model [141], [176]. It classifies both real time errors (active effects) and precursors (latent effects) that lead to accidents. Precursors manifest in a chain over time, from organizational influences to unsafe supervision to preconditions for unsafe acts to the active errors themselves.[24] In a detailed review of HFACS, Stringfellow made two major conclusions: investigators will not be able to consistently classify errors into types, and the method lacks a system-theoretic, design-specific context so that engineers can improve safety in actual systems.

The first issue (inconsistency of findings) is echoed by Dekker. Even with teams of investigators working on accident analyses, results between reports will vary depending on the composition of the team investigating [4]. For HFACS, this fact has been recently confirmed by King [293]. The number of HFACS classification codes was recently reduced from 147 to 109 because the original descriptions for each code were lengthy (paragraph length) but still vague, and investigators were having trouble differentiating between buckets. Even after the change, the correlation of classifications between different groups analyzing the same accidents did not improve very much over the original value of 60 percent.

Regarding the second issue (no system context), without a functional control model of the system in which unsafe behaviors can be identified and documented, it is very difficult to ascertain the hazards within a system by waiting for reportable accidents to occur. HFACS does acknowledge organizational influences, such as policies, procedures, and culture, which are part of its classification structure. However, HFACS assumes mishaps only occur due to a chain of latent (human) failures in the organization and supervision that lead to an active failure by the operator and then a mishap.[25] Furthermore, the mishap-error types are tallied to produce descriptive statistics, and if a certain type of error code is occurring frequently in the field, more training for that type of error is emphasized for all operators in all systems. HFACS thus depends on both analytic reduction and the law of large numbers in order to inform safety knowledge. In doing this, underlying causal scenarios might not come to light within the context of a system's specific design. HFACS and similar methods focus on the human operators and on maximizing their potential to break a mishap chain, relying heavily on hindsight bias to inform its data set [37]. There are many unsafe behaviors that occur in complex systems

[23] At the same time, monetary thresholds for mishap severities were readjusted in the AF regulations.
[24] CRM academics are also based on this concept.
[25] Conversely, in STPA-RC any single developing influence or setting/configuration can directly affect the algorithm of the system's real-time controllers.

due to more than merely failures, and to identify them requires a proactive hazard analysis based on systems' specific designs [182].

*Aircraft Information Program*

AFSEC consolidates the knowledge, policies, and procedures for many of the safety practices mentioned above. They also maintain a common repository for every unique AF mission system that consists of incident reports, mishap-prevention practices, and Military Flight Operations Quality Assurance (MFOQA) information. MFOQA is "the analysis and trending of aircraft flight performance and system data to proactively enhance combat readiness through improvements in operations, maintenance, training, and safety functions…to establish a baseline for normal operations; identify, mitigate, and monitor operational risks while detecting precursors to aviation mishaps; and identify operational inefficiencies" [234], [294, p. 1].

MFOQA serves to aggregate various analyses to identify systemic issues during field operations. Issues that become recurrent could inform updates of system procedures and practices as well as TTPs. MFOQA is one of the facets of a more general practice called the Aircraft Information Program (AIP). Every unique aircraft system in the AF has an AIP database that is managed by AFMC program managers. Besides MFOQA and mishap data, the AIP database includes component integrity, maintenance prognostics, and reliability and maintainability (R&M) data [275], [295], [296]. The AIP and its subset of processes, to my knowledge, do not use a common model of the system with which to put any tracked issues into context.

*Component and Production Quality*

For government systems, a separate quality assurance (QA) requirement exists that is mandated independently by Title 41 of the Federal Code (as opposed to Title 10 which regulates most of the policies discussed up to this point). This requirement mandates timely reporting of product deficiencies, defined as "defects or nonconforming conditions which limit or prohibit the item received from fulfilling its intended purpose" [13, p. 803]. The AF accomplishes this through the acquisition quality program [221], [275], [297]–[299]. Commercial industries use similar QA programs for their product development.

The acquisition quality program catalogs manufacturing defects as well as component defects found during system operations. Defects found during operations are captured in the AF with deficiency reports (DR).[26] Their purpose is to bring operating failure information back to the design and manufacturing process. Those failures do not necessarily have to be determined to affect safety, but those that do are given a higher priority. Both testers and field users may generate DRs; however, field reports are usually given a higher priority for correction unless an issue during test is so crucial that safety for all operations is arguably compromised.

[26] The STAMP organizational control structure example (Figure 2-7 in Chapter 2) includes suggested controls in the design stage to handle manufacturing aspects. Issues found during operations are captured by problem reports in the same diagram.

The DR mindset has its origins in the reductionist, component-based approach to mishap prevention and mission assurance. While they are framed to handle component inadequacies, emergent behaviors that are only explainable by referencing the system design and/or operating philosophy are not easily reported. However, despite being antiquated in format, DRs do offer an advantage as a safety-related practice in that they are by nature specific to the product (and not just a best practice). Test practitioners will sometimes utilize the DR process with a more systemic mindset if they foresee a design issue found during test that might cause emergent unsafe behavior during field use. Documenting those issues comprehensively is often difficult in the absence of a common safety-control model of the system. Instead, a DR of that nature must be authored in a narrative convincing enough for AFMC program managers to understand its systemic implications and prioritize a correction over other reports coming in from the field.

The use of DRs by testers to report systemic problems is a prevalent technique by human-engineering experts, particularly in UV testing.[27] Issues as simple as inappropriate interface-designs and as complex as poor technology-assisted team procedures within a UV GCS can contribute to causal scenarios in the field use of a system. Unfortunately, whether an issue is documented and reported successfully depends on whether the engineer is experienced enough to recognize it, assertive enough to study it (including discussions with unofficial contacts in the field), and willing to devote the resources to writing a DR with a convincing narrative. Human-engineering issues also involve much more than component inadequacies, so DRs are limited by nature. Other safety practices discussed above cover various aspects of human involvement in safety, such as conducting ORM before a sortie, periodic CRM training to shape the operators' mental communications models, and HFACS databases that attempt to classify human behavior.

*Test Safety*

All the other safety practices discussed in this section up to this point have been directly regulated by policies and instructions maintained at or above the AF (Pentagon) level. AF test safety, however, maintains its regulatory practice down at the AFTC level and its goal is to implement the MIL-STD-882 associated management-task of test-and-evaluation risk management [11], [24]. Test safety is a finely applied derivative of AF mishap-prevention policies that takes 882 principles into account as well as operations-safety practices [29], [300].

Although flight testing was a well-established enterprise by the 1970s, the need for test safety was not acknowledged by the AF until an accident at Edwards in 1977 that involved an A-10 aircraft. The design of the test was deemed to have contributed to the incident; minimizing procedures had not been established, and leadership felt they had been

[27] I was given access to DRs for real AF UV systems, but the number of write-ups and their details are not publicly releasable.

unaware of the risk involved.[28] This prompted the creation of new policies that required a hazard analysis be performed by testers on any new projects they receive from AF acquisitions managers. This happened in conjunction with MIL-STD-882 receiving its first revision to version "A" which established the risk matrix and the mandate for command authorities to accept program risks.[29]

The current safety-planning process in AFTC consists of identifying hazards specific to the test, performing a type of PHA (called a test-hazard analysis, or THA), applying risk-reduction efforts, and accepting residual risk before commencing test operations. The heart of test-safety planning is the THA process. The concept of identifying and documenting test hazards is influenced by probabilistic underpinnings, and risk matrices are used to plot the one or multiple hazards found during the planning. Furthermore, the planning process has traditionally assumed the view in Figure 4-2 (a), only looking for hazards in the design and operation of what that view calls the SUT. This tends to focus less on the legacy and test-framework aspects of the entire SDT. Section 4.3 goes into more detail on this and compare traditional and STPA-based approaches in both their philosophy on risk reduction and in the models for their hazard analyses.

AF DT practitioners use "system-safety techniques, prior experience, legacy system research, and overall engineering judgment" to identify hazards and populate the risk matrix [24, p. 10]. Traditional hazard-analysis methods combine likelihood and risk estimates (similar to PHA) with iterative reviewing that leverages extensive aviation and test experience and best practices from technical and operational experts. Both techniques are systematic and have long legacies in the industry, but they do not represent a systemic view. Probabilistic assessments do not often have accurate technical data to support likelihood estimates, and historical statistics and past performance—although informative—are not predictive per se. An expert-review process is also subject to variability due to individuals' educations, experience, and other biases which may result in an incomplete analysis.[30]

One area in which the view of test safety is slowly migrating towards Figure 4-2 (b) is in the region of overlap between test safety and range safety. A test-range safety

---

[28] During a gun-firing test, the flight profile of the aircraft and characteristics of the gun design resulted in secondary ignition of the gun's exhaust gas. Both engines were starved of oxygen because of this shared factor, a type of phenomenon the safety community calls a common-cause accident [301].

[29] The test pilot who was involved in the accident, Francis Gideon, ejected from the aircraft and made a full recovery. Twenty years later he became the AF chief of safety at AFSEC.

[30] One senior safety officer told me that during the safety-plan review, one person can get a bad feeling about the likelihood of a hazard, regardless of what the probabilities say, and affect the overall risk determination. However, these interjections are dependent on who is sitting in the review, and the review board will not always be in agreement. The safety officer gave an example of a board he sat in for a test involving the jettison of newly designed fuel tanks from a fighter aircraft. Both tanks (one under each wing) were supposed to jettison in unison, but the safety officer wanted to assume the worst case that one tank might get stuck, causing an asymmetric flight scenario. His appeal was not considered because the aircraft contractor had PRA results all but guaranteeing the tanks would both drop appropriately. Sure enough, one tank did stick during a real flight test, and the test team had no pre-established minimizing procedures or corrective actions for that scenario. Thankfully, the airplane landed safety.

office was recently stood up in the 412TW to work closely with the other safety offices—shown in the green box in Figure 4-5—to coordinate on test-framework aspects of system testing. This acknowledges that standard range safety practices are often outdated for the new technologies that are being evaluated at Edwards. A senior tester mentioned during fact-finding interviews that test-range safety policy for UVs is somewhere in between, "you don't fly unless you follow every requirement of the RCC documentation," and safety precautions that are "good enough" for the intended purpose.

At least two communication links between a GCS and its UV are required per traditional range guidance, but the test-range-safety office is considering other measures for lost-link scenarios that do not require such a stringent system design constraint.[31] Furthermore, the limitations of the casualty-estimation tools discussed above mean that there is no standardized method to assess requirements for emergency flight termination systems (FTS) [224]. Range-safety officers (RSO) are responsible for assessing UV flight conditions and destroying the vehicle if it exceeds particular boundaries that violate a probabilistically-determined (quantitative) casualty footprint.[32] While the tools might sometimes disagree during test planning, real-time calculation is even more impractical.

The test-range-safety office has attempted to integrate many of the available legacy tools and establish guidelines for how locally trained RSOs should assess the need for FTS activation. They are considering writing a local supplement to the AF range policy to allot for test-specific planning needs, including improvement and agreement of footprint estimators [34]. The acknowledgement that the test stage is unique and requires an all-inclusive view of safety is a good step forward. If UV systems will be performing many of the tasks traditionally managed by MV human pilots, the need for an FTS might be arguably eliminated in some cases as the maturities of the detection and decision-making capabilities of newer systems is more accepted. This view requires a willingness to model the full SDT to include all aspects of the specific design, airspace and range practices, operating practices, and the test framework. It also requires controlling hazards through simple and robust mitigation steps without depending on reliability data.

A senior range operator, during a fact-finding interview, acknowledged that more systemic methods are needed because, "it's interactions that lead to accidents!" Many senior practitioners appreciate the ability to use experience and common sense to contribute to test-safety plan reviews, because it allows them to consider worst-case scenarios to guard against *Murphy's Law*[33] and qualitatively adjust a project's risk. A system-theoretic process

---

[31] For example, they could work with airspace managers to establish some contingency transponder signals the UV could transmit in the event of a lost link; this would allow ATC to work traffic around the UV as it performs a link-recovery flight profile.

[32] FTS is a concept inherited from ballistic missile testing; it is intended to be independent from the system being tested. When applied to UVs, this means the assessment of flight conditions should come from a separate feedback than what the GCS receives, and the termination should not be accomplished through the aircraft components but rather a separate device.

[33] Murphy's Law was a term coined at Edwards. Its meaning has evolved to, "If anything can go wrong, it will" [302, p. 13].

for hazard analysis takes that concept and formally implements it by use of visual-planning tools and a consistent method of analyzing the system for hazard contributions.

#### 4.2.2.3 Airworthiness

While test-safety planning addresses the management of hazards for the test process, testers are ultimately responsible for evaluating the system for its intended fielded purpose and performance. Part of this determination is contributing data to verify the safety of the system as designed and intended for use. Testers can facilitate this in part by using DRs as problem reports to highlight unexpected issues that arise when a system is operated. Testers may also be asked, as part of their evaluation tasks, to contribute verification data to confirm design assumptions that are associated with system safety.[34] Some of these data requirements come from certification requirements.

Many system-certification guidelines and requirements exist in different industries and engineering disciplines. The AF has recently delineated airworthiness as a separately regulated assurance in order to comply with Federal Aviation Administration (FAA) and international standards for flight-system certification [303]. Airworthiness is the "verified and documented capability of an air system configuration to safely attain, sustain, and terminate flight in accordance with the approved aircraft usage and operating limits" [268, p. 2]. Most AF systems must meet various types of certification,[35] and airworthiness applies to flight systems (which fulfill a large portion of the AF mission). While MIL-STD-882 encapsulates the basic process of risk mitigation and the safety-management tasks for the lifecycle of a system, airworthiness prescribes distinct verification standards and methods that must be applied to a system's specific design.

Conceptually, airworthiness tries to ensure system safety by establishing criteria for the design and manufacture of a flight system. It does not, however, analyze how a system is operated. Philosophically, airworthiness only has designers establish the system's operational limits, and users of the system are expected to stay within those limits.[36] The approved certification for any given system is thus valid for a specific configuration and operating envelope. From a systems perspective, airworthiness *is* safety, but it is not systemic and complete. First, it is not broad enough because it does not include operating considerations. Second, the considerations that it does cover (design and manufacture) are governed extensively by PRA.

Within MIL-HDBK-516, *Airworthiness Certification Criteria*, airworthiness is divided into different engineering and management categories that each impose various checks, requirements, and best practices on system design and manufacturing practices. Those partitions include [28]:

- Systems-Engineering Management

[34] This is why a well-specified set of evaluation parameters should include the designer intent and ultimate high-level purpose of the specification(s) [9].
[35] Recent updates to cybersecurity certification, for example, have been proposed by the National Institute of Standards and Technology in the form of best practices [300].
[36] Additionally, airworthiness expects the system to be properly maintained.

  - e.g., R&M, specification documentation
- System-Safety Management
  - e.g., PHA, THA, OSH, hazard tracking
- Aerodynamics and Flight Management (and redundancies)
  - e.g., stability, control, flying and handling qualities
- Subsystems (and redundancies)
  - e.g., hydraulics, environmental control, fuel, fire protection, power, mechanical components
- Diagnostics
  - e.g., built-in tests, sensors for warnings, cautions, and advisories
- Avionics
  - e.g., displays, symbology, normal and failure modes, indications for warnings, cautions, and advisories
- Electrical (and redundancies)
  - e.g., heat and ignition protection
- Electromagnetic/environmental effects
- Computers
  - e.g., computing power, software, autonomy and failure modes
- Maintenance
  - e.g., inspection requirements
- Crew (humans)
  - e.g., escape, life support, lighting and visibility, ergonomics, crash and load survivability, human performance
- Propulsion
- Structures
- Materials

Each of these areas references various additional military specifications and standards that provide further detail. This results in numerous intricate lists of compliance items that must be signed off to certify any new flight system or upgrade. The handbook requires FMEA and similar quantitative probabilistic analyses for the tasks in all the engineering disciplines, including humans and software. The human-engineering standards additionally call for designing a system to minimize workload, maximize situation awareness (SA), improve performance, and minimize human error [122]. While the airworthiness process requires each design to meet various specifications and to use best practices, it lacks the

guidance for how full systems-based hazard analyses might be performed on the specific designs themselves.

Airworthiness certification tasks begin in the design stage and continue into the test stage. This is appropriate because DT should contribute to assessing system safety but is not the sole source of establishing or verifying it. According to a software engineer, during a fact-finding interview, the typical airworthiness certification for a new software capability takes designers six to eight months and 4,000 hours of engineering work to satisfy. For designers, the airworthiness process is considered much more taxing than the process of providing testers data to support their test-safety planning. This may not come as a surprise, based on the current method in which test-safety planning is performed (quick test hazard identification supplemented with expert-practitioner discretion). The use of a system-theoretic process would provide an understanding of the complete system that is being evaluated in order to model the complete SDT. It is concerning that the test-safety process does not always request or have consistent access to more intricate system models and assumptions which designers are already producing for airworthiness.

A recently discontinued document, MIL-HDBK-514, *Operational Safety, Suitability, and Effectiveness*, made an effort to put both DT and airworthiness in the context of complete system safety [279]. It treated the outputs of the design, DT, and even field stage as inputs into airworthiness and—to a larger extent—even the operating safety considerations that airworthiness does not cover. Unfortunately, 514 was difficult to maintain given the vast amount of changes in the various areas addressed within it. Airworthiness is now addressed by several instructions at the AF and AFMC levels and an increasing number of substantiating bulletins which are used as implementation guidance. As airworthiness grew in prominence over the last five years, the AIP, MFOQA, mishap prevention, and incident investigation regulations were also evolving. The attempt to keep 514 updated was abandoned.

Future UV systems will present unproven, immature technologies, particularly in the autonomy and software disciplines. The airworthiness process has and will be challenged to handle autonomous systems, which introduce complexities and emergent behaviors for which traditional certification techniques are inadequate. Currently, there are some airworthiness requirements that must be met by designers before an initial flight clearance is given to testers to begin evaluations. There is the potential that design and simulation data may not be sufficient to gain clearance without actual flight-test data to verify assumptions and update predictions. This obviously presents a dilemma. To avoid such a problem, product safety should be framed in a system-theoretic viewpoint. This begins with designers building a safety-control structure and conducting a hazard analysis they can deliver to testers, so at least a common framework is being used across the lifecycle to answer questions of top-level safety.

Furthermore, safety certifications should not be solely based on PRA if the systems framework described above exists. Airworthiness is an emergent property of a system, and reductionist calculations are not enough to certify it. Future work could propose a hazard-analysis management process for AF acquisitions that is truly a systems-engineering tool, not just a checklist of design requirements. It would also include operating practices, which

the current airworthiness process does not currently include. Good systems engineering provides the framework onto which the discipline expertise can be applied. It avoids specification solutions without traceability to top-level properties.[37] Airworthiness (verified and validated through design, test, and fielding) should evolve to be true system safety, capable of being treated with systems techniques. Through proper systems engineering, such an approach would preclude current practices that develop specific solutions that result from ignoring top-level system requirements.

### 4.2.3 Explicit Influences

Much of the discussion in the preceding sections has referenced various policies and regulations. Appendix B presents and discusses the basic layout and appearance of an EIM constructed for the 412TW, developed during a thorough policy review. This section puts the referenced documents from the previous sections in perspective and demonstrate the ease in which an EIM can be used to discuss policy and how it flows to the front-line test activities. Instead of bibliographical citations, the alphanumeric designations of the documents are included parenthetically as they are mentioned. The figures in Appendix B should be referenced along with this section.

At the top of the EIM is Title 10 of the Federal Code, which governs the activities of the armed forces and DOD. Directly below Title 10, there are many regulations and directives (many more than are shown on the map), but the two that serve as the primary influence for all the policies governing AF DT are mapped (DODD 5000.1 and DODD 3200.15). At this level there are already many standards and guides that influence all military practices, and some of them are bundled in boxes to the left (e.g., MIL-STD-882E, DODAF, and various system-safety and test-and-evaluation guides). Next, below the top defense directives are policies describing major offices that manage defense acquisition. The Undersecretary for Acquisition, Technology, and Logistics (DODD 5134.01) manages the practices of the design and DT stages of product development, while the Director of Operational Test and Evaluation implements initial field evaluations (DODD 5141.02).

Below the authority of the Undersecretary for Acquisition, Technology, and Logistics are many applicable policy sets. All the general safety and occupational-health policies in the military flow to the left side of the map (DODD 4715.01). Airspace and coordination with FAA activities flow almost down the center (DODD 5030.19). Range (DODD3200.11), test management (DODD 5105.71), airworthiness (DODD 5030.61), and lifecycle management (DODI 5134.16) make up the remainder of the major acquisition regions. On the far right of the EIM is policy for user-requirements generation (CJCSI 3170.01). In the AF it is the using commands that work with AFMC to generate future system needs.

The groupings described above are farther elaborated by AF policy directives. Starting with “Military Legal Affairs” (AFPD 51-5) and going to the right, a progression can be followed through the groupings of safety, operations, airfield/range management,

[37] It also avoids specifications for human and software behavior, which cannot be easily modeled quantitatively.

test and evaluation, airworthiness, system lifecycle management, and requirements generation. Incident investigations are governed by two sources that cover public accident investigations (AFI 51-503) as well as privileged safety investigations (AFI 91-204). Occupational safety (AFI 91-203) and mishap prevention (AFI 91-202) round out the main safety programs, while closely related is risk management (AFI 90-802). From mishap prevention flows test safety (AFTCI 91-203).

Aircrew operations and procedures occupy the next major section of the EIM. Of note in this area is CRM policy (AFI 11-290), general flight rules (AFI 11-202v3), and test-specific flight rules (AFI 11-2FTv3). System-specific policies are also shown on the map in the form of RQ-4 Globalhawk procedures (AFI 11-2RQ-4v3) and a box symbolizing the Globalhawk tech manual(s) and TTPs.[38] Test policies are supplemented by various academic and training programs. The year-long TPS curriculum is used to train flight-test aircrew, and policy and training also exists for mission-control room procedures (EAFBI 99-108).

Airfield (AFI 13-204v3), airspace (AFI 13-201), and range (AFI 13-212) regulations all affect general Edwards flying policies (EAFB 13-100). The flying policies also implement flight-test policies from the left on the map, and they are supplemented by numerous forms of airspace guidance written for both local practitioners and visitors to the Edwards ranges (all in a box above and to the right of 13-100 that includes R-2508/2515 airspace charts, user guides, and local area orientation training).

Test and evaluation policy (AFI 99-103) governs both DT and field evaluations in the AF, and it flows down to test policies in AFMC (AFMCI 99-103) that focus on DT. Continuing down, AFTC implements these policies into test planning (EAFBI 99-101), test conduct (EAFBI 99-105), and test reporting (EAFBI 99-103). Those are the three segments of the product test stage. Airworthiness (AFI 62-601) policies closely complement testing, and MIL-HDBK-516C provides detailed guidance for airworthiness. Farther to the right are various policies that manage the AIP (AFI 63-133, AFI 63-140), as well as applicable standards and handbooks.

Lifecycle management (AFI 63-101) covers many lower policies. Most notably are the quality assurance (AFMCI 63-501) and DR policies (AFMCI 63-510). The detailed DR process is outlined in a technical order (AFTO 00-35D-54), and AFTC implements its own process (EAFBI 99-224) for meeting QA requirements. This process is a form of test reporting and is shown as such on the map.

Test control and conduct (EAFBI 99-105) is the lower focus of the map, and the area that makes this EIM specific to the 412TW test enterprise—and more narrowly, the test operations of the 412TW. While upper portions of the map will look similar to the test wing at Eglin AFB, the lower explicit influences are more specific to the way in which policy and procedures are channeled at Edwards. Below 99-105, the operations group policies and instructions reside. Temporary restrictions, called flight-crew information files (FCIF), are collected at this level from various sources and published, serving as broadly

[38] The EIM was constructed while I was performing research with the Globalhawk flight-test squadron, so that system was used as an exemplar.

applicable settings/configurations. Below, the flight-test squadron has its own policy and instructions. Both the group and the squadron have a standard format for briefing test missions and monthly or quarterly commander interest or emphasis items. Finally, each test project is marked by various goals, planned flight-test techniques, test points and maneuvers, and safety-planning considerations. During test, systems have certain temporary or permanent operating limitations, watch items (usually anomalies), and active DR entries. All of these influences are explicit and can be labeled as standards, rules, or settings/configurations as determined by practitioners.

MIL-STD-882 affects many items on the map as has been discussed previously. Although formal safety policies occupy the left side of the EIM, safety practices are implemented in various forms, and they come from various disparate sources of authority. Any effort to consolidate the methodologies for assessing system safety needs to be aware of all these policies and practices as they evolve and relate to one another.

The EIM can be used to manage and understand the propagation of knowledge gained and created from valuable lessons over time. In 2009, an F-22 aircraft was involved in a fatal accident at Edwards [304]. The circumstances involved a test setup that required the airplane to intentionally dive toward the ground at very high angles. After the incident, some resultant policy changes initially came in the form of FCIFs limiting dive angles for all testing. Eventually, these limitations were moved to test-operations policies (AFI 11-2FTv3). After a thorough analysis by test practitioners over the next year, a new method for managing the collision potential with the ground called Time Safety Margin (TSM) was developed. In short, it is a calculation of how much time (in seconds) an aircraft has to pull out of any specific dive geometry to preclude an imminent ground impact. TSM can be calculated for all planned test maneuvers for a given project, and it can even be computed and displayed in real time. Once the method was approved, 11-2FTv3 was updated to replace dive-angle restrictions with TSM minimums, and 99-105 was also updated to make TSM a required briefing item and data entry on aircrew flight cards.

TSM is an example of practitioners deciding on what matters in their work processes [153]. By codifying the concept and method of calculation of this parameter, testers can use it as a new behavior state in their mental models during operations. During test planning, they can discuss target values for TSM and question the necessity and value of the planned maneuvers for product evaluation. Explicit knowledge such as this and sanctioned definitions, methods, and training can be easily traced on the EIM. Additionally, inherent within some of the explicit influences are tacit influences developed through academics, training, and experience. School programs, classes, simulators, proficiency exercises, and recurring briefing standards supplement the documents and policies on the EIM.

The EIM is a product that should be managed centrally but made available to all practitioners to use personally to understand and shape test policies. It can be used to train newcomers to the enterprise on the explicit policies that affect their work domain. It can

also be used by leadership for a more efficient communication to the organization of new policies and updates. [39]

### 4.2.4 Incorporating Test into STAMP Hierarchical Control Models

The system designers have preconceptions of how a new system should behave, how it will be operated, and the type of environment it will encounter. However, these preconceptions are based on unverified engineering models, predicted field-user techniques, and uncertain estimates of field-environment characteristics. If products were to be fielded for real-world evaluation immediately after initial design and manufacture, it could result in unexpected or unacceptable performance, ill-knowledge of physical operating limitations, and in some cases unsafe behavior inherent by design.

When the acquisition timeline permits, the transition of a system from design to regular field use is accomplished through a range of assessment activities. DT and field evaluation anchor the two ends of this spectrum. Sometimes their efforts combine and/or overlap to expedite the fielding of a system. However, establishing a logical divide between them makes sense in order to highlight and manage the goals that are predominant in each.

As discussed in Chapter 2, the goal of field evaluation is to "demonstrate, under as operationally realistic conditions as possible and practical, that systems are operationally effective, suitable, and capable of meeting the user's requirements," while the goal of DT (accomplished earlier in the lifecycle) can be summarized by the following [219]:

- Specification compliance within a representative environment
- Risk reduction (for technology capabilities, safety, performance, etc.)

Field evaluation validates that the emergent behavior of the system satisfies its mission requirements. The operationally realistic conditions required for this validation include the system itself as it is configured to be fielded, the TTPs used to operate it, and the work environment. Traditionally, DT primarily verifies component behavior and only requires representative conditions, meaning there is some ability to control the certainty of the phenomena that the components of the system are exposed to during test operations.

Designers have a notion for how components should behave so that the system will perform its mission well and without accidents. One way they document this is with specifications that initial testers use to evaluate the components. A second way is to document incomplete or missing predictions as knowledge gaps that require further risk reduction in DT. Any design assumptions that were not explored with DT activities are carried over to field evaluation, particularly with regard to aspects of regular operations and maintenance (O&M) practices and long-term quality and reliability of components.

Safety of complex systems does not reside merely in component behaviors, however [5]. Without a recognized control hierarchy that describes how functional entities

[39] I received a very positive response when building and demonstrating the EIM as a tool at Edwards. A flight-test squadron commander was so impressed he briefed it to the operations group leadership. The practitioners in charge of maintaining the local standards repository were also impressed, and I am currently working with their chief to investigate the feasibility of implementing such a tool at Edwards.

of a system work to detect and regulate potentially hazardous scenarios, the evaluation of the system will not adequately address hazard mitigation. This is true for the activities of DT, field evaluation, or any combination thereof. Because safety is irreducible to component behavior, designers should pass safety assumptions to testers in a system-theoretic format [9].

Having the knowledge of a system's design intent—to include preconceptions of fielding conditions and assumptions for emergent (not just component) behavior—can give DT professionals the ability to report problems that would be more expensive and timely to correct later in the system lifecycle. Furthermore, by using a systems-based model and analysis technique that is geared toward documenting emergent properties, many contributions to hazards could be anticipated before test operations even begin.

Leveson's generalized example of a STAMP organizational control structure as it existed prior to this thesis is shown in Figure 2-7 in Chapter 2. It has now been updated in this thesis to include a test stage (DT) that captures the basic relationships and flows of information as a system transitions from design to initial field evaluation. The updated diagram is shown in Figure 4-7. The test stage is unique in that a large, consistent framework exists to create sanitized conditions to isolate system and component behaviors and to perform data measurements during operations. Developmental products enter and exit this framework, necessitating a professional enterprise that can combine its consistent corporate expertise and specialized assets with the various and unique unproven designs as they arrive at the test stage to begin evaluation.

The test stage in Figure 4-7 replaces the region in Figure 2-7 that was labeled as maintenance and evolution. The sparse horizontal decomposition between development and fielding that this former region contributed is now thoroughly fortified in the new diagram with information requirements that are discussed in more detail below. Maintenance and operations are incorporated in both the test and field stages. Similar to Leveson's diagram, each depicted stage contains an operating process, as indicated by the outlined box at the bottom.

The three-stage paradigm used by the DOD provided the template used to update Leveson's diagram. The management entities that oversee each stage remain generically independent in Figure 4-7. The management structure depicted is a general example that can take a more specific form depending on each unique industry and how its product development is structured. In the AF, for example, the first two stages of the product lifecycle report to AFMC while the third reports to the using commands. In some cases, one company might be solely responsible for a design while another independent group stewards the testing. Different entities might be involved in regulating and fielding products.

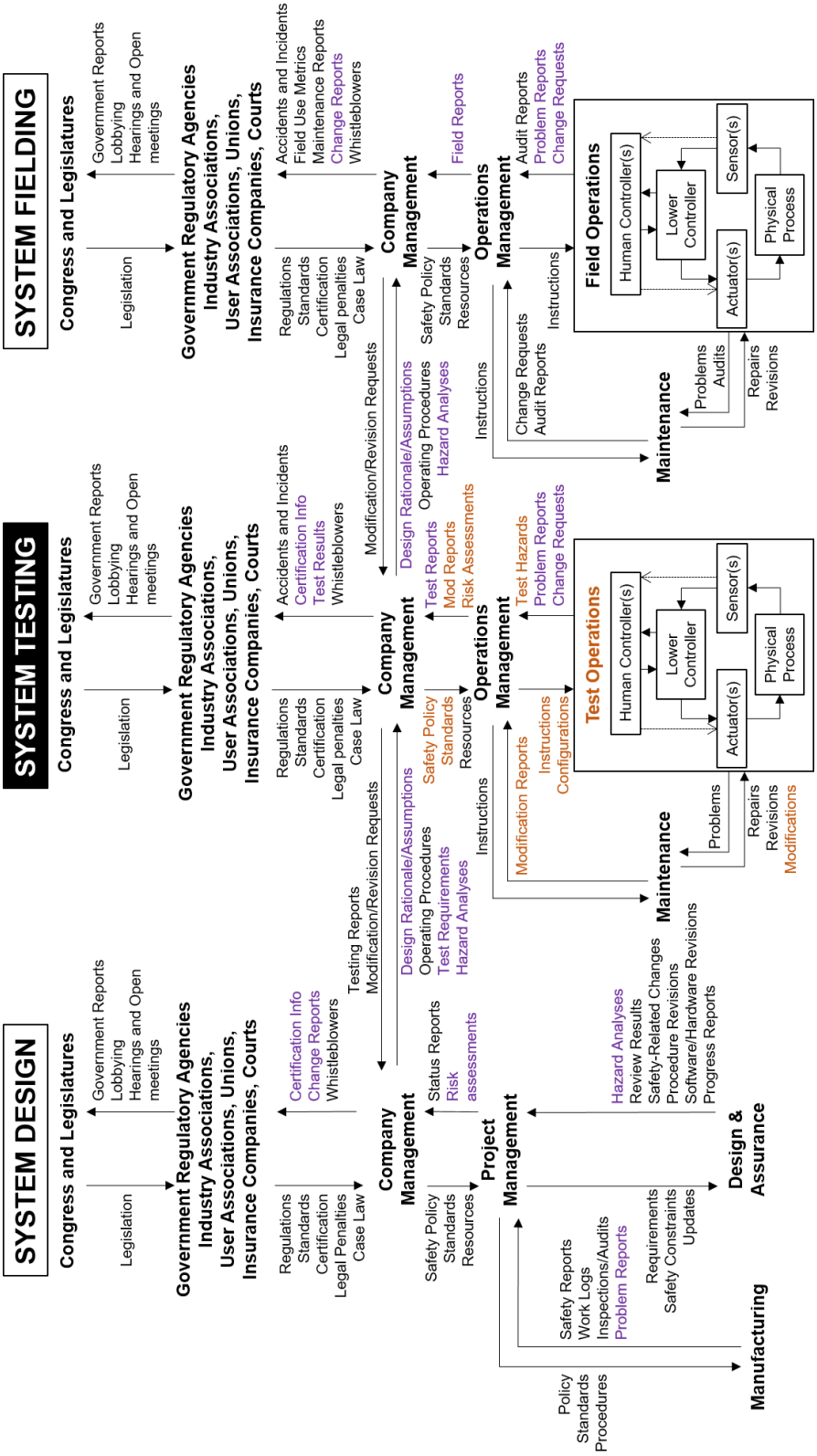



Figure 4-7. Updated Organizational Control Structure Example

What the new organizational control structure example adds to Rasmussen and Leveson's contributions are specific safety considerations brought about by recognizing DT during development. For control models to successfully encompass these three stages of product development in an organization, the stakeholders must accept that a physical product is not the only thing that passes from one stage to the next. What is vital for system safety to be assured consistently and efficiently is for traceable product documentation to be communicated between stages as well. This includes control models and analyses, maintained in a common format that can be challenged and updated by practitioners along the lifecycle stages. Furthermore, these models are systemic and updated as appropriate to encompass the aspects of the particular sociotechnical enterprise at each stage of product development.

The two goals of safety during DT are to determine the safety of the system as designed and intended for use and to ensure the safety of the testing process itself (i.e., test safety). Figure 4-7 highlights the controls and communications that support each of these goals in purple and orange, respectively. The non-emphasized connections are important for safety, but the purpose of directing attention to the highlighted items is to underscore the information and activities that the test stage shares with the other stages as well as manages on its own.

System safety begins with designers performing STPA and preliminary risk assessments. Those activities should establish the common STAMP control structure that will be used by testers and field users for the remainder of the life of the system. This model and the corresponding analysis results are passed to the testers, along with the rationale and intentions of the design and the envisioned field-use concepts. Also communicated are test requirements. They include component specifications and risk-reduction requirements for more complex behavior that designers could not accurately forecast. The DT stage, with a full understanding of the justification for both component behavior and system behavior requirements, reports on their evaluations in order to update design assumptions as well as field-use concepts. An updated control model and risk assessment are then passed to field users so the system may be exercised in a representative environment. By this stage, most inherent design issues should have been corrected, and field evaluators can focus on the system's ability to meet mission requirements as designed and under real-world TTPs.

Both design and DT contribute to fielding certification, as indicated by the information being fed up to higher agencies in Figure 4-7. A lack of an approved certification should not be a reason to limit the activities of DT; testers exist to assume the added risk of operating unproven systems in order that they may *inform* certification. Certification requirements and achievements should be included in the information that is shared horizontally between developmental stages. When necessary, testers may report problems with components, the design itself, or the operating procedures. Using a STAMP control structure gives discipline engineers in the test stage a framework to document their findings. Problems are reported similarly in the field; they can be put into the context of the functional control model so that designers and testers may understand why the problems matter at the system level.

Table 4-2. Safety Information Across Stages

| | Design | Test | Field |
|---|---|---|---|
| **Produce** | Initial Control Model<br>Initial STPA<br>Component Specifications<br>Knowledge Gaps<br>Certification Data | SDT Control Model<br>Design Deficiencies<br>Component Deficiencies | Fielded-System Control Model |
| **Update** | | STPA (test framework)<br>Verified Specifications<br>Knowledge Gaps<br>Certification Data | STPA (evolution)<br>Design Deficiencies<br>Component Deficiencies<br>Knowledge Gaps |
| **STPA: New Assumptions** | Component Behaviors<br>Emergent Behaviors<br>Field Operations Concept<br>Field Maintainability Concept | Test Framework Behavior<br>Test Limitations | |
| **STPA: Renewed Assumptions** | | Emergent Behaviors<br>Field Operations Concept<br>Field Maintainability Concept | Future Operations Concept<br>Future Maintainability Concept |

Test safety is the other safety goal of the test stage and is necessary not only because of the unproven aspects of the system being evaluated, but also because of the techniques, configurations, and approaches used for confirming design specifications and updating assumptions. The test framework contains unique aspects including apparatuses, targets, instrumentation, policies, techniques, and hardware/software modifications to legacy items. When these aspects are included in the model of the system, new control actions and causal scenarios can contribute to hazards. Before testing, DT practitioners must update the control model received from designers to incorporate the test framework, thus forming the SDT. Figure 4-2 (b) demonstrates the SDT as a non-structured representation; an example of structured representation is shown in the next section.

Table 4-2 summarizes the key information that is created and updated throughout the lifecycle stages when product development traces safety systemically. The first two rows of the table present this information. Besides creating certification data, it is important for designers to document test requirements (e.g., specifications, knowledge gaps) so that DT can verify measurable behaviors to contribute more certification data and address assumptions. This documentation is more straightforward when it can be tied to a safety-control structure and an STPA analysis that the designers originate. The last two rows of the table show the assumptions that can be addressed within a shared STPA analysis that is updated throughout the development stages. In design, the hazard analysis initially contains many assumptions regarding system behaviors, operating techniques, and maintainability.

Once testers update the designers' control model to refine it into the SDT, they must always question which aspects of the SDT are different from the conceptual fielded system,

including which additional items are needed for the test mission that would not be found in the field, and which features of the field cannot be replicated. Testers must also modify the STPA analysis to include behavior and limitations of the test framework and the techniques under which the system will be exercised. As specifications are verified and gaps in knowledge addressed, assumptions about behavior may be updated, as well as the operating and maintenance concepts. Subsequently, those in the field continue to keep the STPA analysis updated as future evolutions of operating practices and in-place system modifications occur. Extended-use data in the field contributes to more knowledge gaps being addressed and initial maintainability techniques being improved.

From a systems perspective, DT is a unique stage not just because of its ability to sanitize conditions for mere component verifications, but because the test framework combines with the IBE and legacy items to form an entire system. Modeling the SDT requires extensive knowledge of the test framework. For flight test, the test framework includes airspace management and control, range and telemetry, mission control rooms, instrumentation of the IBE, and flight-test squadron personnel and their operating procedures. Each test project is unique. Furthermore, within each project, every operating phase is new and different as the system is evaluated and knowledge gaps are addressed. Periodic updates to the hazard analysis are recommended as more is learned about the IBE and emergent system behavior. Conversely in field operations, phases (e.g., briefing, operations, and maintenance) are more consistent and predictable, and systems evolve gradually.

Being prepared in the test stage to accept a new product and appropriately model the SDT can benefit from understanding the sociotechnical backbone providing the corporate expertise and assets of the test framework.[40] In their work on risk-management modeling and organizational human factors, Dulac and Stringfellow, respectively, created and refined inclusion criteria for both the system design/development and system fielding portions of Leveson's generalized example of the organizational control structure [21], [188]. The following list supplements those criteria by adding considerations for engineers wishing to begin creating control models for the test stage.

INCLUSION CRITERIA FOR SYSTEM TESTING:

1) Is the actor/entity/component responsible for or involved in defining requirements, criteria, and metrics for test enterprise capabilities and test-project schedule priority?
2) Is the actor/entity/component capable of influencing the allocation of resources (e.g. funding, staffing) throughout the enterprise?
3) Is the actor/entity/component capable of hiring/firing controllers within the system?

[40] At Edwards, this would include the 412TW, its safety offices, the operations group, engineering group, maintenance group, support group, flight test squadrons, discipline-engineering offices, instrumentation shops, range support and operations teams, and airspace and airfield managers that were introduced earlier.

4) Is the actor/entity/component responsible for enforcing schedule pressure, budgets, and/or resource requirements (especially safety requirements) for systems during test?
5) Is the actor/entity/component responsible for defining test standards, practices, and processes (especially safety-related standards and processes)? If so, does it have enforcement power?
6) Is the actor/entity/component capable of changing the requirements, standards, procedures, or waivers for test operations or influencing others to do so?
7) Does the actor/entity/component perform a significant amount of work on activities such as safety analyses, system maintenance, system integration, and/or quality assurance?
8) Is the actor/entity/component responsible for, or heavily involved in system modifications for test?
9) Is the actor/entity/component responsible for, or heavily involved in, system certification renewal or review?
10) Does the actor/entity/component have the authority to request a delay or stop in production when problems arise?
11) Is the actor/entity/component an important contractor of the system, providing a significant portion of the system hardware or technical and operating personnel?
12) Would the actor/entity/component be impacted in the event of an accident?

For DT to treat safety as a systems problem, the SDT should be analyzed using the top-down approach STPA provides. The stakeholders of the test enterprise should identify accidents and hazards, noting if any of them are different than those that designers used for their initial STPA documentation. This identification is important because accidents and hazards guide the hazard analysis. If they are different in the DT stage, different UCAs and causal scenarios might be identified due solely to that change. The other reason that UCAs and causal scenarios will be different in the test-stage analysis is due to the presence of the test framework within the control structure.

Accidents and hazards unique to the test stage should be highlighted along with test-framework items within the safety-control structure. The UCAs and causal scenarios that arise that are related to those aspects can be acknowledged to be test-specific. They may then be distinguished as such in the planning documents with any chosen method of emphasis. In this way, *test-safety planning is a natural product of the greater system-safety analysis* of the SDT. Test safety is thus a special case of system safety. Furthermore, because the overall analysis is systemic (STPA), the test-safety findings will be as well (e.g., not just component failures).

Testers, if circumstances permit (or require it), may also begin updating the STPA documentation of a product with causal scenarios based on field-use assumptions. The test stage includes socio-organizational aspects, and controllers in the SDT are affected by the influences that flow down through the test enterprise. Test practitioners, when thinking

ahead to the safety of the fielded system, should consider how policies and other explicit influences might differ between the test and field stage as well as how staffing and manning affect the knowledge and skill of the personnel who will be operating within that system in the field. These scenarios may also be emphasized in a distinctive manner in the STPA documentation similarly to how test-specific scenarios were highlighted. That is one way to renew the field operating assumptions per Table 4-2.

Testers may even be able to use actual field techniques to begin assessing more emergent system behaviors earlier. While this might be useful when those behaviors involve only (or mostly) machine components, there are limits to how realistically the testers can themselves emulate the *human* (and coupled human-machine) behavior that would be observed in the field. Part of the TPS curriculum trains test pilots how to behave like less-experienced pilots when evaluating the handling qualities of an aircraft.[41] It is difficult to find many other cases in which test operators can specifically train to behave like novices. While manual control is a human-system integration (HSI) subject that has been quantified and engineered over decades of research, more complex HSI phenomena involving supervision, decision-making, and communications are not so easily discretized into modular techniques, if at all.

If test pilots cannot emulate field behavior for anything but the most straightforward performance tasks, inherent system safety cannot be completely validated only by experienced testers operating within the SDT. It is still important to have a common safety-control model shared by the lifecycle stages with assumptions updated appropriately until field users can confirm them. The thorough system-theoretic hazard analysis that begins during design should include human-controller analyses to document assumptions for field behavior. STPA-RC is an example of a method that engineers can use early in design to identify causal scenarios involving complex human behavior. As the system progresses toward fielding, those scenarios would be revisited as the system is assessed in order to renew the assumptions about operator behavior.

Often, schedule and budget pressures necessitate more focus on test safety during the test stage. Without a system-theoretic control model, local test planning assumes a reductionist approach in which safety mitigations are too often applied haphazardly or as an afterthought. The next section discusses this more and shows that STPA, if need be, can still be used to focus on test-specific UCAs and scenarios for a single operating phase in the absence of the opportunity to analyze the entire SDT. It is not ideal, but it still leverages the power of STPA to find systemic causal scenarios that go beyond component failures.

Figure 4-7 visualizes a test stage in order to consider a complete system-development lifecycle. It captures the basic controls and communications necessary for tracking products, their control models, and their design and use assumptions through the stages of development. Problems and unanticipated behavior can be documented in a

[41] Several methods exist for amplifying the stress and response gains of a pilot. One involves very determined tracking of a single target in space. Another involves the avoidance of simulated hazardous boundaries in the airspace that steadily shrink around the flight path. These techniques encourage test-pilots who are normally very experienced and anticipatory controllers to abandon their expert manual-control techniques in order to rate the qualities of an airplane as they would exist for those less experienced in flying [106].

system context by anyone in any of the stages. The same control model and hazard analysis is shared across lifecycle stages to manage the general case of system safety (including certification) as well as the special case of test safety, and any additional regulatory requirements such as QA can be put into its context.

Stringfellow said, "No matter how it is implemented, a force that resists migration to high risk within the organization is crucial to system safety" [21, p. 83]. In my experience with the fact-finding interviews at Edwards, I determined that testers are always looking for the opportunity to comment when a system does not achieve its requirements and discuss how it could be better. Test practitioners gravitate towards experiences that are technically challenging and require a keen management of uncertainty. The ability to consistently report not only on test safety but the safety of the product itself, when documented in an organized and systems-based format, will take advantage of testers' attention to detail in the most efficient way possible.

The next section demonstrates test-safety planning using STPA. Although sometimes a hazard analysis on an entire SDT may not be possible given schedule constraints, meaningful findings can come out of performing the hazard analysis on enough of the system to capture the IBE and the methods used to evaluate it.

## 4.3 Test Safety Planning

With the system-theoretic perspective of testing described above, a demonstration now follows using a real project's test-safety planning to demonstrate the utility of using a common system framework to guide hazard analysis. This allows for not only a useful analysis of the scenarios that contribute to hazards due to testing, but also a contribution to inherent system-safety documentation for the lifecycle documentation of the product. Tracing and updating assumptions about the product safety allows decision makers at every stage to prioritize hazard mitigation strategies as programmatic tradeoffs.

The lower-center portion of the updated organizational control structure example in Figure 4-7 (containing orange text) emphasizes the activities involved with test-safety planning and documentation. Planning for test safety is a purposeful activity unique to the test stage of a product. The traditional method and format of test-safety planning are introduced next, followed by a proposed method and format for an STPA-based approach. I had the opportunity to draft a notional STPA-based test-safety plan for a real flight-test project that was conducted at Edwards during the period of this research. Additionally, I collected viewpoints from eight AFTC test practitioners to gain subjective insight from these professionals on the intelligibility, informativeness, and implementability of the STPA-based technique for test safety.

A project containing features of autonomy and/or human-machine interactions was desired as it would present an ideal case[42] to assess a new type of hazard analysis, applicable to the emerging challenges at AFTC. An interesting opportunity in the UV

[42] Practitioners with whom I worked found it quite interesting that the AFTC test-planning method itself was being evaluated.

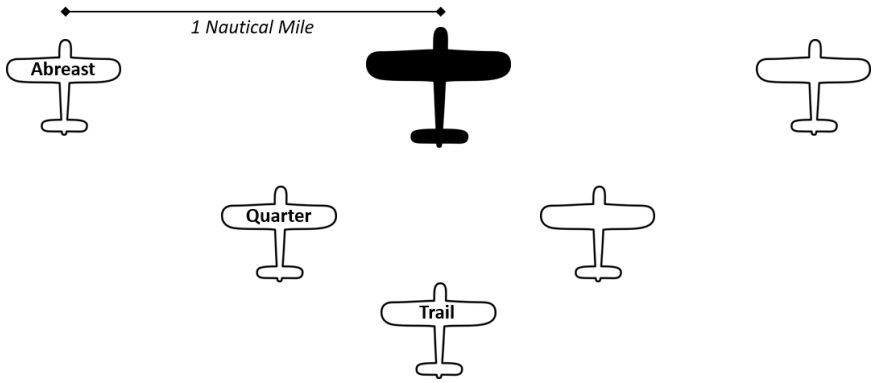



Figure 4-8. Wingman Formation Positions

domain was found, in the form of a proof-of-concept system evaluated at TPS. The project is described and then used as an example to compare the traditional and STPA-based methods. The traditional safety-planning format is discussed first along with the results for the UV project. Then the newly developed STPA-based safety-planning format is introduced with its results for the UV project. Comparisons between the two methods and their results follow. The UV project is discussed in minimal necessary detail to avoid disclosing sensitive project data and AF TTPs. Neither the traditional safety plan nor the notional STPA-based safety plan for the wingman project is entirely reproduced in this thesis.

### 4.3.1 Description of Flight Test Project

The test project, which was planned and executed by TPS students during the thesis-research period, was designed to demonstrate a proof-of-concept for an autonomous wingman [305]. In its basic premise, a flight formation of two aircraft would have a manned platform as the formation lead. A single aircraft would fly as the wingman in a geometrically-defined relative position off the lead. The wingman would fly autonomously, and the possible positions could vary between wing abreast (roughly one nautical mile directly to the left or right of lead), trail (several thousand feet behind lead), or an intermediate angle and distance resulting in a rear-quarter position. See Figure 4-8 (lead is black and wingman is white).

For testing, the lead and wingman were both the same type of aircraft and both had onboard pilots. Each platform had a modern, fly-by-wire digital flight computer that converted human-pilot inputs into control-surface deflections. There was no direct input from the pilots to the air environment, and aerodynamic feedback was only received through visual (cockpit window) and vestibular physics.[43] Additionally, both flight computers contained a proven, proprietary software algorithm that allowed for automatic collision avoidance. If this mode were to be enabled on either aircraft, its flight computer would continuously calculate an avoidance cone between the ownship and the other

[43] A safety-control structure is shown later. It would benefit the reader to envision the functional relationships in the system, based only on the description here, before continuing on.

aircraft. If certain position and velocity closure criteria were met, the flight computer would execute an avoidance maneuver without human-pilot input.

Use of the collision-avoidance mode itself on either aircraft was not part of the planned testing; however, the wingman aircraft's algorithm was modified to use the same time, space, and position information (TSPI) about itself and the lead aircraft that was normally available to the collision-avoidance algorithm. Using those data, the wingman could autonomously fly in formation with respect to the lead aircraft. This ability effectively made the wingman a UV surrogate. It had two human pilots onboard that could supervise the aircraft without providing manual input while the autonomous "pilot" flew the aircraft.

The wingman flight computer obtained TSPI about its ownship from its onboard global positioning and inertial navigation system as well as its air-data computer. The computer also used this information to project its TSPI several seconds into the future. The current and projected TSPI of the other (lead) aircraft was made available to the wingman via radio transmitted datalink between the aircraft. This datalink was made possible by a pair of communication pods; one was physically attached under the wing of each aircraft. The data transmission parameters were new and proprietary, so the pods themselves were modified beyond their legacy configurations.[44]

The test maneuvers for the project included various station-keeping tasks in the different formation positions while the lead aircraft flew straight and made turns. The autonomous algorithm was evaluated on how well it anticipated lead's maneuvering and compensated to maintain the appropriate transitional and final positions in the formation per AF TTPs. Lead could request the wingman to transition between abreast, quarter, and trail, as well as switch from left to right and vice versa. Those types of requests could be communicated in one of two ways: lead could send a push-button message through the datalink requesting the change, or lead could perform a predetermined acute maneuver (such as a wing rock) that would be understood by the wingman algorithm by interpreting lead's TSPI data. The method of station-change request that the autonomous pilot would obey was selectable by the human pilots inside the UV-surrogate wingman.

When the wingman was abreast, lead could also request the entire formation to change its flight direction. Lead could only request these formation turns through acute maneuvers, not push-button messages. The maneuvers were more abrupt than the ones used to request station changes (e.g., a full wing flash instead of a gradual wing rock). Once initiated, the formation turns involved more prolonged geometric coordination between the aircraft (as dictated by TTPs), with several different possible transitional maneuvers that determined the resultant heading of the formation.

---

[44] The pods were modified to support the appropriate data packets and rates for the purposes of enabling the autonomous flight capabilities of the experimental flight computer in the wingman aircraft. The required fidelity of TSPI for flight-formation maneuvering is higher than that which can be provided by Automatic Dependent Surveillance-Broadcast (ADS-B), the real-world intership navigation network used as an example at the end of the previous chapter.

In addition to formation testing, lead could command the wingman—through push-button messages—to perform several types of flight maneuvers on its own. These included loitering in a holding pattern, following a preplanned navigation route, and rejoining with lead to assume a formation station. Each of these behaviors were evaluated to satisfy a test objective. Because the UV was a surrogate, its on-board pilots could evaluate each individual maneuver while the surrogate was far away from the lead aircraft. Additionally, pilots were encouraged to provide comments about the behavior of the system to satisfy a final test objective aimed at gauging operational utility.

The two pilots onboard the UV-surrogate wingman had distinct responsibilities. When the surrogate aircraft was in test mode, the evaluator pilot would manually fly and set up each test point via direct flight-control inputs, preconfigure the autonomous pilot to the correct mode for evaluation via push buttons, and then engage the autonomy via push button. He could also enter an altitude offset between the wingman and the lead aircraft to provide a safety buffer for formation and rejoin maneuvering. Once autonomous, the evaluator pilot could not control the wingman aircraft with his flight controls. He would have to disengage the autonomous pilot using the same control panel and then revert to manual flight controls. The safety pilot, however, could instantly disengage test mode and take control of the aircraft from the evaluator pilot (disengaging both the evaluator and autonomous pilot control) merely by touching the safety-pilot's own flight controls. This override mode would give the safety pilot sole (manual) control of the wingman aircraft.[45]

TPS test projects are not as large in scale as the evaluations being performed in the formal flight-test squadrons at the 412TW. However, they go through the same test-planning segment as any other test project and thus appropriately represent the test wing paradigm. In the planning segment, the technical and safety planning are accomplished and documented and the system is modified and configured as necessary by instrumentation and maintenance teams to prepare for the test conduct segment. Once test conduct begins, typical repeating phases apply: mission (sortie) planning, briefing, test operations, debriefing, and between-sortie maintenance.

### 4.3.2 Traditional Planning

Traditionally, safety planning in the 412TW is a form of PHA [11]. Test planners in the flight-test squadron, a combination of test aircrew and engineering-group discipline experts, conduct a hazard analysis and document it in the safety plan with a proposed risk level. The plan is reviewed by a board of senior engineers and operators from outside the squadron and updated if necessary. The plan is then approved by a pre-determined level of management in the 412TW that is dependent on the initially estimated risk level. See Chung for details on the planning phases [249].

The documented safety plan complements the documented technical plan. The technical plan is typically written by the engineering-group personnel assigned within the flight test squadron and then approved by senior staff in the engineering-group home office.

---

[45] This is a standard crew configuration for flight-control evaluations in fly-by-wire aircraft. Typically the evaluator pilot sits in the left seat and the safety pilot in the right seat.

The wing safety office encourages engineers to consider safety aspects as they write the technical plan and to write the safety plan concurrently with the technical plan, but in most cases the technical plan is already going through approval as the safety plan is drafted. Often many components of the technical plan are summarized or repeated in the safety plan, because a different group of reviewers staffs the safety plan. Once the safety plan is approved, it becomes an annex to the technical plan, and the test planning is complete.

Beyond the technical summary and general safety considerations contained in the first portion of the test-safety plan, the next portion forms the core of the document. It summarizes the hazard analysis and contains the safety implementation measures for the project. Performing a hazard analysis meets the intentions of the historical system-safety movement, and the AFTC guidance is derived from AF mishap-prevention guidelines [24], [29]. Those guidelines describe hazards as real or potential conditions that are precursors to mishaps.[46] The only hazards that are required to be identified in AFTC are called *test hazards*. They are defined as hazards that are unique to the IBE[47] and the test framework but not present in the normal operational hazards associated with the system or environment. Furthermore, over the years many safety plans have identified some mishaps (e.g., midair collision) as a hazard. The local safety-training curriculum[48] concedes this is allowable if planners do not foresee any actions that can be taken between the recognition of the condition and the mishap, or if the first recognizable phenomenon is the mishap itself. This caveat results in an inconsistency in the justifications and conventions used to identify hazards (more in Section 4.3.4).

Figure 4-9 summarizes the chain-of-events model that forms the basis for the traditional test-hazard analysis, or THA. Test hazards are considered to be different than non-test hazards—called *general hazards* here for the sake of discussion. Typically, test hazards are identified for documentation anytime planners foresee a causal chain involving aspects of the testing (i.e., IBE and/or test framework) that leads to a mishap. As mentioned previously, the test hazard might be the last detectable condition before the mishap or the mishap itself.

Once a list of test hazards has been generated, each one is examined. No standard is mandated here, and rarely are visual-planning products such as event trees or fishbone diagrams even included in the documentation. Each THA is documented on a standardized worksheet that is written in text (example shown in Section 4.3.2.2). First, the causes of a test hazard are identified. There can be multiple causes, and the guidance recommends looking for design inadequacies, component failures/malfunctions, procedural deficiencies, personnel error, and environmental conditions.

Next, one or multiple mishaps are identified (e.g., injury/death), and the severity category of each is determined, typically following published injury and cost criteria. Following that step, *minimizing procedures* (MP) are identified that might break the chain between the cause and the test hazard. MPs are either directives or considerations and can

[46] This is similar to the STAMP definition of a hazard, but it tends to encourage a chain-of-events mentality for the analysis. The AF definition of a mishap, on the other hand, is almost identical to STAMP.
[47] As discussed earlier, the IBE is called the SUT in traditional language.
[48] *Test Safety Training Academics*, found in the 412TW EIM.

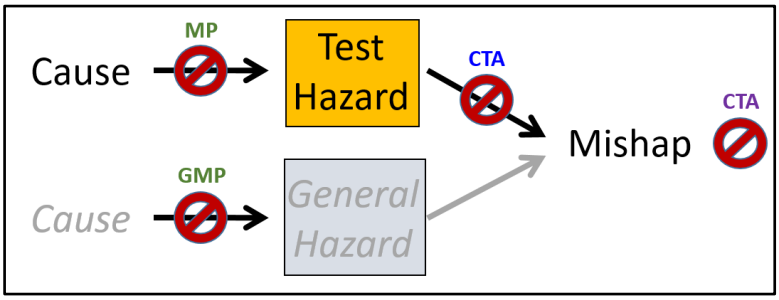



Figure 4-9. Traditional Test Safety Model

be mandated to occur before or during the test mission. Each one is tied to one or more of the documented causes. After MPs are identified, *corrective actions* (CTA) are listed. CTAs are intended to break the chain after the hazard. Because sometimes the hazard *is* the mishap, a CTA might be documented that needs to be performed after a mishap to reduce the effect or severity of the mishap. MPs and CTAs are intended to reduce the probability of the mishap. The planners then estimate a residual risk[49] based on the assumption that MPs and CTAs will be applied appropriately during test conduct. The risk level is qualitatively determined by using a risk matrix, like a PHA.

Documented separately from the THAs are *general minimizing procedures* (GMP). Those are stand-alone statements used to address miscellaneous operational restrictions, system limitations, global requirements for the test project, and parameter monitoring. Like MPs, GMPs are either directives or considerations and can be mandated to occur before or during the test mission. A GMP may sometimes be an identical statement to an MP within a THA worksheet, as MPs tend to repeat when there are many THAs in the safety plan. GMPs are not generated via the same level of rigor as MPs, as there is no test hazard that prompts them. They are akin to practices that minimize general hazards, although general hazards are never actually acknowledged in the planning. There is often discord among planners as to what belongs in a THA and what belongs in a GMP, and the safety office must assist in judging those matters as the plan is being drafted. Unlike THAs, the GMP chain of events does not identify general hazards or CTAs. In Figure 4-9 this omission is indicated by a grayed-out path.

The result of the THA and GMP documentation is a finalized safety plan and a recommended overall project risk, which is submitted to a board of senior practitioners for review. They consider the THA worksheets and contribute their own expert perspectives to determine the final recommended risk level before the plan. Due to variability in manpower scheduling and availability, the panel will consist of a different set of senior practitioners for every project. Although an accident model is assumed (chain-of-events), there are no system or design diagrams or general visual-planning tools available to discuss the justification for the hazard analysis the planners performed. Everything is summarized verbally with the THA worksheets and an overall risk matrix that plots all the test hazards. The ability to consider all the mechanisms for mishaps relies solely on what the planners

[49] Section 4.3.4 discusses how STAMP and the traditional method compare in their risk-reduction strategies.

thought of during their event-chain analysis and the knowledge available from the group of reviewers with their variable set of experiences to draw upon for critiquing the plan.

#### 4.3.2.1 Format of the Traditional Safety Plan

The typical structure of a traditional safety-planning document is presented here. Note that it is divided into two main sections; the first mostly summarizes the project background and the test strategy found in the (separate) technical plan, while the second discusses specific safety-implementation measures and safety mitigations [244]:

SECTION I – PROJECT DESCRIPTION

1) Summary of Changes
   - *Used to track updates any time the safety plan is revised*
2) Background
3) Mishap Responsibilities
   - *Short statement indicating who is responsible for investigating a mishap and by what policy (usually defers to AFI 91-204)*
4) Test Objective(s)
5) Test Item Description
   - System Under Test
   - Test Facilities
   - Test Instrumentation
6) System Maturity / Readiness to Test
7) Predicted / Expected Results
8) Types of Tests

SECTION II: SAFETY IMPLEMENTATION

1) Qualification and Training
   - Aircrew
   - Control Room Personnel
2) Test Article Restrictions
3) Special Considerations
   - *Any additional remarks that generally apply to the test project*
4) General Minimizing Procedures (GMP)
5) Test Hazard Analysis (THA) Worksheet(s)
   - Title of Hazard
   - Residual Risk (shown on the matrix)

- Cause(s)
- Effect (Mishap)
- Minimizing Procedures (MP)
- Corrective Actions (CTA)
- Remarks

#### 4.3.2.2 Traditional Test Safety Mitigations

The following is reproduced from Section II of the TPS students' traditional test-safety plan for the wingman project, with changes or paraphrasing done where necessary to mask sensitive information [306]:

*1) Qualification and Training*

Lead Aircraft - Before piloting the lead aircraft, the pilots on the test team require checkout flights in the aircraft type. One formation checkout flight is required. These checkout flights will occur per the TPS local instructions.

Photo/Chase Aircraft - The pilot of the photo/chase aircraft will have completed the photo/chase upgrade syllabus. The photo/chase will be conducted only on points that have already been successfully demonstrated. The photo/chase aircraft will be positioned to the outside of the formation for all maneuvers. The photo/chase pilot will not be a TPS student.

Control Room - All flights will be monitored in the control room by project members not flying and designated aircraft engineering contractors. The purpose for the control room is to provide real time data collection, feedback, and troubleshooting.

*2) Test Article Restrictions*

With the surrogate test mode engaged, the UV surrogate is limited to [a specific proportion] of the normal aircraft design limit load, and the maximum operating airspeed is [*removed*].[50] The tests have been designed in order to maintain the surrogate aircraft within this envelope. All test points will be flown within the current modified-aircraft guidance.

*3) Special Considerations*

The safety pilot in the surrogate UV has the ability to immediately disengage the test mode (which also turns off the autonomous pilot). Should the need arise, the safety pilot will disengage the test mode and maneuver the aircraft appropriately in override mode.

[50] "Removed" denotes technical details that are proprietary.

*4) General Minimizing Procedures (GMP)*

1. There is a radiation hazard for ground personnel when the datalink pod has power. Maintenance personnel will be briefed before each flight by the aircrew, adhere to the safety distance, and follow all applicable pod handling guidance.
2. All testing will be conducted in day visual meteorological conditions (VMC).[51]
3. The minimum altitude for all test points is 5,000 ft. above ground level (AGL).
4. The air-collision algorithm will be placed in standby on both aircraft and verified prior to formation maneuvering.
5. Specifics for photo chase missions:
   - Photo chase will only be executed on test points that have already been completed on a previous sortie and assessed to have no objectionable qualities, based on pilot comments and engineering analysis.
   - A test point will be terminated if the photo chase loses sight of either aircraft during formation maneuvering.
   - Photo chase will primarily base its positioning off of the lead aircraft. The intent is for photo chase to remain on the outside and/or aft of the formation (keeping the lead aircraft in the middle of the formation). At no point shall photo chase close within 100 ft. of either aircraft.

*5) Test Hazard Analysis Worksheet (1 of 1)*

Title: Mid-Air Collision During Formation Maneuvers

Residual Risk: Severity Category I / Improbable (see Figure 4-10)

Causes:

1. Aircrew error
2. Error in autonomous pilot software
3. Loss of visual

Effect: Death and/or system loss

Minimizing Procedures (MP):

1. (1, 2, 3)[52] All aircrew will be pre-briefed on all maneuvers, areas of concern, responsibilities and terminology in the flight briefing. The emphasis shall be on "blind" communications, deconfliction responsibility, and use of the aircraft transponders for distance finding. Only one distance-finding source is required for the execution of these test points. (See Remark 1)

[51] Day VMC means between sunrise and sunset, without any weather or cloud activity that limits visibility, obstructs clear airspace, or otherwise necessitates the use of cockpit navigational instruments alone to navigate the airspace.

[52] The parenthetical entries are references to the cause(s) that the MP is intended to mitigate.

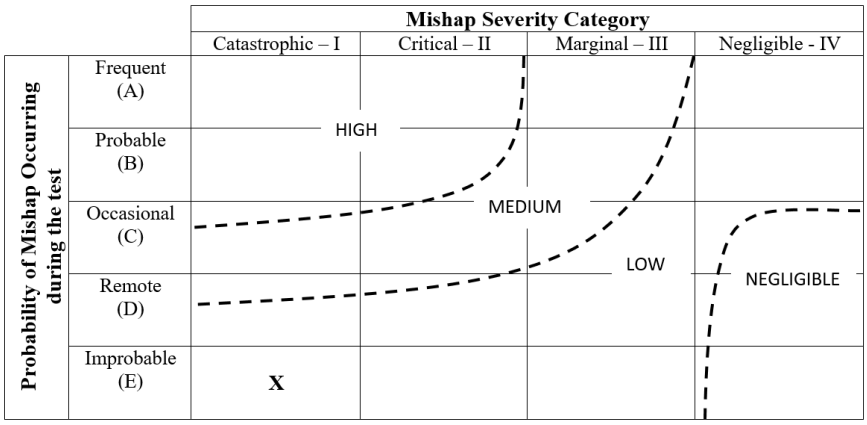



Figure 4-10. Wingman THA Risk Matrix [306]

2. (2) All formation maneuvers shall be executed with sufficient lateral and/or vertical separation to provide room for either pilot to safely maneuver and avoid the other aircraft.
3. (3) "Altitude offset" in the wingman flight-control computer will be set no less than 200 feet below the lead aircraft when in formation.
4. (3) At least one aircraft will always be visual of the other. If either aircraft goes "blind", standard communications shall be used to maintain positive deconfliction until regaining a "visual."

*The following MPs (5–8) apply to rejoin test points:*

5. (3) If neither of the aircraft gain sight of each other by 2 nautical miles, the point will be terminated and the safety pilot in the wingman aircraft will disengage the surrogate test mode, maintain current altitude, and begin coordinating a rejoin via radio communication with the lead aircraft.
6. (3) Minimum separation between aircraft during all rejoin test points is 500 ft.
7. (1) Prior to the first break up and rejoin, while in fingertip formation the two aircraft will ensure their altimeters read the same altitude. This will help mitigate any possible altimeter errors based off of the same barometric pressure setting.
8. (1, 3) To ensure the VISTA does not exceed the 500 ft. altitude buffer, 600 ft. will be entered vas the flight-control "altitude offset" entry. This will account for any slight altitude variations (e.g. ±20 ft.) while maneuvering.

Corrective Actions (CTA):

1. Execute appropriate emergency procedures.

Remarks:

1. The other indication/sensor available in the lead or wingman aircraft to determine distance between the aircraft is the radar.

2. In accordance with TPS instructions, a current and qualified TPS instructor pilot will be in the formation.

### 4.3.3 STPA Based Planning

Because a test stage is newly incorporated into the STAMP view of a system lifecycle in this thesis, and because test-safety planning is a unique activity of the test stage, I developed an STPA-based approach to test-safety planning. The new approach is presented in this section. It is discussed as having a fundamentally different accident model than the traditional approach. A proposed safety-plan format follows, and the wingman project once again is used to demonstrate aspects of the hazard analysis and planning document.

One of the advantages of STPA as a hazard-analysis technique is that it enforces proper systems engineering, which includes good requirements definition and system modeling. STPA is top-down. Accidents and hazards should be well defined and identified within the enterprise. Safety constraints and the safety-control structure serve as the basis for any product-specific hazard analysis conducted by test planners. The SDT includes all aspects of the product evaluation, including the IBE, legacy components, and the test framework. These aspects also include socio-organizational functions of the test enterprise.

The safety-control structure of the SDT is important to have as a common visual-planning tool for test practitioners. Building the control structure does not happen immediately or without great deliberation among stakeholders and engineers. The aspects of the SDT that include the actual product being evaluated should already have a control structure provided by designers with their STPA results. Testers build the control-structure for the SDT by incorporating the test framework. The test enterprise should establish the standards and guidance for how common aspects of the organization and test framework should appear in an SDT control structure. Those common aspects, such as ATC, range assets, and control rooms should be implemented similarly for all individual test projects. The enterprise should also perform STPA on these common aspects (e.g., hazard analysis on the common use of a dune buggy on the range as a mock target for all test projects that need ground targets).

STPA can find contributions to hazards anywhere in the control hierarchy. Schedule constraints will probably preclude test planners from analyzing the SDT starting from its highest levels (e.g., AFTC commander level). Instead, the hazard analysis may be abbreviated to focus on the IBE and the test infrastructure during an operating phase.[53] By adjusting the scope of a project's test-safety planning as such, it is possible to focus on a correspondingly appropriate volume of the safety-control structure. When STPA is performed, the resulting planning document is framed to summarize the mitigations the test practitioners must apply *before* and *during* operations. This is a case where it is useful to have a technique for identifying influences, such as with STPA-RC.

Figure 4-11 summarizes the system-theoretic foundation for a test-safety hazard analysis using STPA. It is a complementary diagram to that in Figure 4-9 for the traditional analysis. It emphasizes that the hazards are the guide to the safety planning. It depicts the

[53] It may also include any other phases during test conduct, as desired.

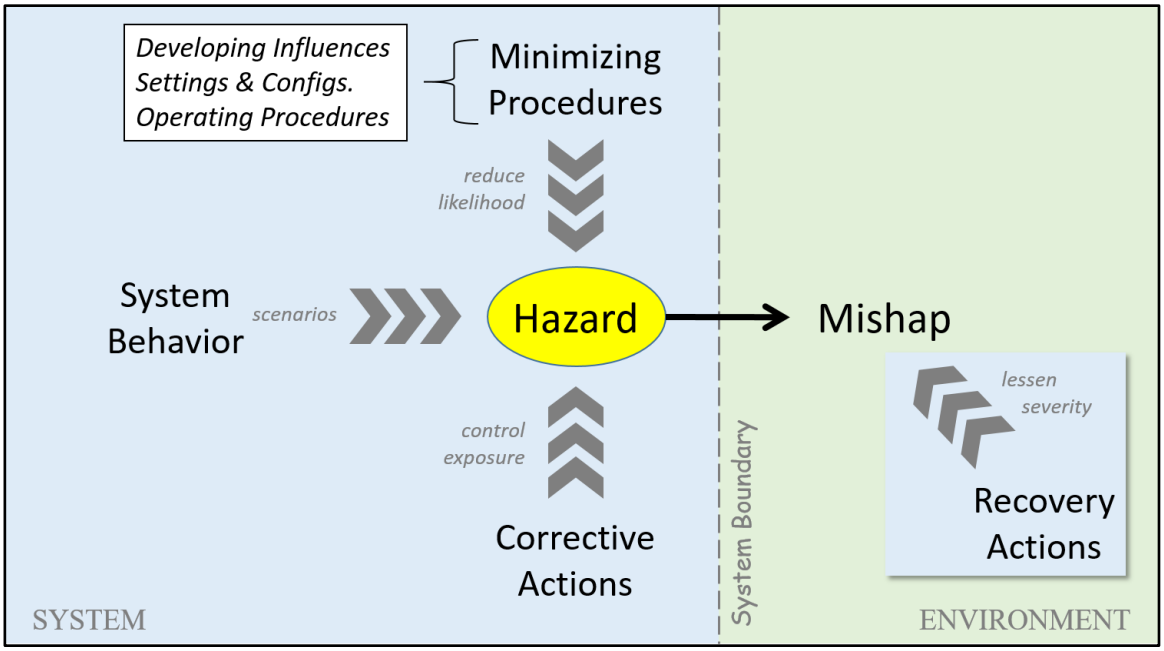



Figure 4-11. STPA Test Safety Model

SDT in blue and the environment in green, along with the system boundary they share. Everything in the system is under the control of the system's owners, while everything in the environment is not. A hazard is shown to lead to a mishap if a particular set of environmental conditions allow it. Because the mishap is on the right of the figure, it cannot be directly controlled. Only the hazard(s) can. In the STPA-based approach, hazards are typically *not* specific to the design or testing of the product. This differs from the THA approach, in which the purpose of the planning is specifically to identify test-unique hazards. With STPA, the hazards are defined based only on the domain, industry, or technology sector the test enterprise supports.[54]

The other parts of Figure 4-11 are arranged to stress that this is not a chain-of-events model. Three areas are related to the hazard(s) in the figure via chevrons. The chevrons should be interpreted as "this factors into", not "this causes". *System behavior* represents all controls and behaviors that are possible within the laws of physics, including those of humans and software. Systems behaviors contribute to hazards in the form of UCAs and the causal scenarios that lead to the UCAs. STPA performed during safety planning looks at system behavior to identify these UCAs and their causal scenarios. *Minimizing procedures* and *corrective actions* are terms borrowed from the traditional format. However, here MPs and CTAs do not break chains of events. Instead, they serve as controls in the system to mitigate the existence of the hazard(s) either before or after the hazard(s) occur.

[54] Unless the test is for a brand new technology that cannot be generalized by other existing hazards.

MPs mitigate hazards by specifying certain controls before and during test operations in order to reduce the likelihood of the hazard(s). Developing influences and settings/configurations have been defined in the previous chapter and begin to occur at some point before the operating phase.[55] Operating procedures are controls that occur during the phase. CTAs mitigate hazards by controlling the exposure of (or trying to remove) the hazard once it occurs and is detected. Nothing can prevent the hazard from becoming a mishap if the environmental conditions allow it. Although the traditional view allows mishaps (e.g., midair collision) to be called hazards, the STPA view does not. A hazard might be difficult or impossible to recognize before it becomes a mishap, or there may not be any CTAs available to mitigate the hazard once it occurs. In these cases, the focus is on MPs. Furthermore, CTAs are no longer affiliated with mishaps like they were in the traditional view. Instead, *recovery actions* (RA) have been defined to be controls and procedures that attempt to lessen the severity of a mishap if it occurs.

Ultimately, safety mitigations (MPs, CTAs, RAs) can be seen as intentional constraints that must be enforced in the SDT, which is a system defined by the nature of the IBE and the way it is being evaluated. Like the traditional safety plan, the STPA-based safety plan must document mitigations in an organized manner that is accessible to reviewers and approvers. Expert opinion is still valued during the review process, but it is aided by the safety-control structure as a visual-planning tool and the ability to trace UCAs and mitigations to it. Event-chains are no longer the basis for the safety planning, although the STAMP model by nature still allows for the identification of appropriate causal scenarios that involve component failures. Visual planning helps refine assumptions. The act of planning this way helps practitioners take ownership of the analysis and maintain a common picture through the interactions among planners [12].

The ideal sequence for an STPA-planning process would look like the following:

a) Receive safety-control structure and hazard analysis from designers modeling the system as intended for fielding, including assumptions and known limitations

b) Modify the safety-control structure to model the SDT for a test-operating phase

c) Establish the accidents and hazards that will factor into the analysis[56]

d) Perform STPA on the new control relationships that exist due to the updated safety-control structure

e) Revisit design-stage UCAs for new contexts that exist due to the updated safety-control structure (particularly the test framework)

f) Redesign the SDT if possible/practical to eliminate hazards

g) Identify MPs, CTAs, and RAs

h) Document the safety plan

---

[55] Developing influences and settings/configurations contribute to both UCA causal scenarios (as demonstrated in the previous chapter) and to mitigating procedures here.

[56] More discussion in Section 4.3.3.2.

i) Expert review and management approval of the plan

The next sections go over the proposed format of an STPA-based test-safety plan document and present the safety mitigations it documented for the wingman project. In constructing the wingman safety plan, I was not aware of every possible aspect of the design, and I did not understand the mechanizations of the autonomy or the surrogate aircraft in depth. I was given and took advantage of access to the same past safety plans from older similar projects as the TPS students that wrote the wingman project's traditional safety plan. I was not involved in writing the technical plan for the wingman project like the students were. I reviewed the technical plan and also viewed Section I (project description) of the traditional safety plan to ensure I was assuming the same basic level of technical understanding before performing STPA. I did not read Section II[57] of the traditional safety document until I had completed the STPA-based document so that I could make a fair comparison. The UCAs for the wingman project I identified are not shown in this thesis as the information on them is project sensitive.[58]

It should be re-emphasized that for a systems view of testing to be effective, a hazard analysis would notionally first be performed by the designers before the system goes to the test stage. Test-stage practitioners do not have the time and expertise to properly identify all UCAs and causal scenarios inherent in the IBE. This lack of in-depth system knowledge does not mean that developmental testers are not concerned with system safety as well as test safety. Testers should be able to begin making assessments about how a system would behave in the field in order to better write the test reports and problem reports. In projects involving autonomy-capable systems, some of the objectives of the evaluations might even be to determine inherent system safety.

#### *4.3.3.1 Proposed Format for the STPA Based Safety Plan*

The format proposed here for an STPA-based test-safety plan draws from some of the structure of the traditional AFTC safety plan.[59] Testers developed the traditional format over four decades of experience in documenting important test-project information to stakeholders. A safety plan must be intelligible and implementable even when trying to convey a large amount of information. The STPA-based format attempts to preserve the core structure of an already successful communication format while making significant modifications to implement the systems view. The resulting guidance is applicable to any industry domain, not just AFTC.

The STPA-based safety plan has three main sections instead of the two in the traditional format. The reason is that an additional section has been added to the front to summarize planning information for decision-makers (e.g., reviewers, approvers). This section (I) also contains the summaries of any revisions as well as a general remarks area to capture comments and special considerations that were formerly distributed throughout other portions of the traditional format (such as within THAs).

---

[57] This is the section of the traditional safety plan containing safety mitigations including GMPs and THAs.
[58] The control structure and variable reference are still shown in Section 4.3.3.3.
[59] I also incorporated feedback I received at Edwards as I was developing the STPA format.

Section II of the safety plan is now the project description, and some of its aspects with respect to the wingman project are discussed in Section 4.3.3.3. It includes programmatic information about the product such as designer, sponsors, contractors, and responsible organizations for investigating mishaps. Technical test objectives are repeated from the technical plan, and the SDT is described, including the display of the safety-control structure and a discussion of system modes. Required and desired assets for the test conduct are listed. A very important part of this section adds system limitations. It is here that testers must document their knowledge of design assumptions, and make references to the safety-control structure where necessary. Understanding the limitations as well as the strategy for building up the test approach to relieve limitations (when applicable) is crucial to establishing appropriate restrictions when safety mitigations are considered later.

Section III of the plan is now the safety implementation. Its structure is noticeably modified from that of the traditional plan. The goal for Section III is to provide intelligibility while allowing for traceability. Section III discusses the actual hazard analysis from the top-down perspective, beginning by listing the accidents and hazards of concern for the project. From there it walks through each test maneuver or evaluation technique and discusses the aspect of the IBE being examined,[60] the description of the evaluation technique, expected results, and the identified UCAs that apply to it.[61] This format puts UCAs in the context of the technical plan. Safety mitigations are then presented.

The mitigations based on the STPA analysis of the wingman project are discussed in Section 4.3.3.4. Mitigations in the STPA-based safety plan are quite different than in the traditional format. Restrictions are covered first (based on the system limitations identified earlier in the document), then minimizing procedures,[62] corrective actions, and recovery actions. Each MP is traceable to one or multiple hazards (and/or UCAs or causal scenarios), each hazard has one or more CTAs, and each mishap has one or more RAs. The traceability of MPs is difficult to convey merely with a written document.[63] The recommended method is to number the MPs in the safety-plan document and then provide decision-makers access to the tables or databases that contain the MPs listed against UCAs (which themselves already trace to hazards). This method would give reader the option to reference the hazard analysis without cluttering the written plan and making it unintelligible. The fact that MPs might mitigate multiple UCAs or causal scenarios is not a new feature introduced by the new safety-planning format; in the traditional format, MPs can mitigate multiple causes, and identical MPs are commonly found among different THAs in the same safety plan. However, the new format allows MPs to be consolidated into single unique statements that do not repeat.

---

[60] To include emergent properties of the system itself (e.g., utility).

[61] The hazard analysis (STPA Step 1 and 2) is performed by planners and the results are maintained in tables or databases. However, the raw results should not be presented in anything but an appendix to the safety plan.

[62] Training qualifications are absorbed into MPs.

[63] I originally considered parenthetical comments or footnotes for each MP in the STPA-based safety plan. However, the feedback I received from practitioners at Edwards what that this type of formatting made the document ungainly and difficult to comprehend, especially if an MP traced to a large number of UCAs.

The STPA-based safety-plan format is outlined below in its entirety. Annotations in the footnotes accompany the outline to expound some areas and provide example language from the wingman-project safety plan.

SECTION I – PLANNING SUMMARY

1) Summary of Changes
   - *Used to track updates any time the safety plan is revised*
2) Overview of findings[64]
   - Total number of test objectives and number of evaluation methods
   - Total number of hazards

     Total number of UCAs
   - Total number of total minimizing procedures (MP)
     - Subtotal number of developing influences
     - Subtotal number of settings/configurations
     - Subtotal number of operating procedures
   - Total number of hazard corrective actions (CTA)
   - Total number of mishap recovery actions (RA)
3) Remarks
   - *Any planning, system configuration, or mitigation tasks that have been accomplished and are required to be reported, emphasized, or signed off by reviewers/approvers go here*[65]
   - *Any general remarks or special considerations (not directive in nature) to bring attention to unique aspects of the project should also go here, especially any clarifying remarks/considerations that are made within the document*

SECTION II – PROJECT DESCRIPTION

1) Background
2) Mishap Responsibilities
   - *Short statement indicating who is responsible for investigating a mishap and by what policy (usually defers to AFI 91-204)*
3) Test Objective(s)[66]

---

[64] Wingman: 4 test objectives, [*removed*] evaluations/maneuvers, 4 hazards, 392 UCAs, 46 MPs (14 developing influences, 10 settings/configurations, 22 operating procedures), 8 CTAs, 7 RAs.
[65] Wingman: "All test maneuvers have been practiced in the simulator by all project test pilots."
[66] Wingman: a) Formation flying ([#] methods); b) Formation rejoins ([#] methods); c) Mission Elements (route following and loiter, [#] methods); d) Operational utility (1 method – practitioner comments).

   - *For each, include the number of evaluation methods*

4) Description of System During Test (SDT)[67]
   - Safety Control Structure
   - Item Being Evaluated (IBE)
   - Legacy Items and Modifications/Configurations
   - Test Facilities
   - Test Instrumentation
   - Control Discussion
     - System Modes
     - Required and Desired Assets[68,69]
5) System Maturity / Limitations / Readiness to Test[70]
6) Predicted / Expected Results

SECTION III: SAFETY IMPLEMENTATION

1) Safety Requirements
   - Accidents[71]
   - Hazards
   - General Safety Responsibilities[72,73]
2) Types of Tests[74]
   - *Go through each maneuver/evaluation one at a time: include the aspect of the IBE being examined, description of the maneuver, expected results, and the UCAs identified that apply to it*

---

[67] The traditional format called this section the test item description.
[68] Required and desired assets can be deemed so for technical and/or safety reasons.
[69] Wingman: There were quite a few required assets including functional transponders on both aircraft and clear visual sight between aircraft; a desired asset was the test-control room which could provide "backup for ranging setups, test-point setup and cadence, data collection, and engineering troubleshooting."
[70] Wingman: Hardware and performance limitations were discussed for, among other things, the surrogate aircraft's flight-control configurations and aerodynamic envelope, the autonomy algorithm's flying and station-keeping logic, and the transmit and receive capabilities of the modified datalink pods (which were considered part of the test framework, not IBE).
[71] The accidents and hazards for the wingman example are discussed in Section 4.3.3.2 of the thesis.
[72] This part provides the option to clarify general responsibility, authority and accountability of controllers.
[73] Wingman: "For formation and rejoin objectives, lead aircraft and ATC have responsibility for traffic deconfliction, between the formation and other aircraft, while the surrogate pilots have responsibility for ensuring formation spacing. For route-following and loiter objectives, the surrogate pilots have responsibility for ensuring traffic and terrain deconfliction."
[74] This part is where the STPA Step 1 and 2 findings are put into context of the test activities.

- General operation, airspace/range transitions, test-point setups and transitions[75]

3) Safety Mitigations[76]
   - System Notes and Restrictions[77]
   - Testing Restrictions
   - MP: Developing Influences
     - Technical and Safety Planning
     - Training and Qualification
     - Flight and Test Manuals
   - MP: Settings/Configurations
     - Test Card Requirements
     - Briefing Requirements
     - Instrumentation and Item Configurations
     - Special O&M Considerations
     - Special ORM / Physiological Considerations
   - MP: Operating Procedures
   - Hazard Corrective Actions (CTA)
   - Mishap Recovery Actions (RA)

#### *4.3.3.2 Accidents and Hazards with Example*

Accidents and hazards should be established before beginning the hazard analysis. For a review of accidents and hazards and how they fit into the STAMP view, reference Chapter 2. The accidents (mishaps) that are of concern to stakeholders in the test stage will typically at least include any accidents that field users from the same industry-domain are concerned with. There may also be additional accidents in the test stage due to the location of the testing or because of types of mission losses that are exclusive to the test enterprise. During my research I met with 412TW stakeholders and formed a suggested list of the accidents important to Edwards leadership. This list then served as the basis for the wingman-project STPA analysis:

A1: Ground personnel are killed or injured

A2: Ground assets are damaged or destroyed

---

[75] After each type of test maneuver is discussed, this last part documents the same considerations as above but for mission transitions before, after, and between test points, and general operation of the SDT. This ensures safety is implemented for the entire sortie.

[76] This part of the safety plan for the wingman example is reproduced in Section 4.3.3.4 of the thesis.

[77] System and testing restrictions establish general criteria for the actual mitigations (MPs, CTAs, RAs) and are based on the limitations of the SDT.

A3: Flight personnel are killed or injured

A4: Flight assets are damaged or destroyed

A5: Asset enters prohibited airspace or range

A6: Test data are lost or destroyed

It would be at the discretion of 412TW leadership to codify their accident list, as it will change very rarely. The list can also be ordered. It is impossible to reduce the likelihood of all mishaps to zero, so prioritizing accidents by severity helps stakeholders decide which safety mitigations on a project to implement given other constraints (such as cost and schedule).

Before commencing with safety planning, test-project planners must work with enterprise safety management to agree on the system boundaries of the product being tested. Within those boundaries, hazards can exist. As discussed earlier, hazards are defined based only on the domain, industry, or technology sector the test enterprise supports; they are rarely unique to the design or testing of a specific product. I recommend that safety management maintain a standardized repository of hazards that testers may choose from to construct a list appropriate to the test project.[78] For demonstration purposes, I drafted the following notional list of hazards for use in the wingman example, with parenthetical traceability to the 412TW accidents:

H1: Aircraft violates minimum separation distance to other flying objects (defined by range and closure rate per domain policies) (A1-A4, A6)

H2: Aircraft violates terrain closure limits (defined as a trajectory and energy state which, uncorrected, could result in a ground collision) (A1-A4, A6)

H3: Aircraft departs aerodynamically stable flight (A1-A4)

H4: Aircraft exits allowable testing area (A5, A6)

Because hazards can also be written as safety constraints, that list follows for reference:

SC1: Aircraft must not violate minimum separation distance to other flying objects

SC2: Aircraft must not violate terrain closure limits

SC3: Aircraft must not depart aerodynamically stable flight

SC4: Aircraft must not exit allowable testing area(s)

#### 4.3.3.3 Example Safety Control Structure

Figure 4-2 represents the change in mindset from a system *under* test to a system *during* test. The SDT contains the physical product, personnel, use philosophy, and organizational

[78] Currently, the 412TW safety office maintains a repository of past THAs and encourages planners to reuse them to maintain some level of consistency in the archives. It is rare for the worksheets, however, to remain consistent as every test team makes minor changes and suggestions to every recycled THA. A shorter, simpler list of hazards would be much easier to keep standardized. Not all hazards will be applied for all test projects (e.g., "human exposure to laser energy" would not apply in types of tests that do not involve lasers).

structure of the test stage, including anything under the control of designers, practitioners, and organizational stakeholders (including airspace/ranges, support assets, and policies). The IBE aspects of the system are represented in purple, while the test framework is orange. Section 4.2.4 provided inclusion criteria that test-stage practitioners should consider when modeling their enterprise with a hierarchical control structure. The control structure also establishes common aspects of the test framework like airspace, control room, and range entities.

When the test stage receives a new product, testers must work with designers to understand the configuration of the IBE, including its interface with legacy items that sometimes must be modified to accept the IBE (e.g., a new experimental flight-control system on an old airplane). Then test planners must develop the strategy for testing the system, which includes revising the designers' safety-control structure to incorporate the test framework to model the product-specific SDT. Ideally, the SDT control structure would include all the phases within test conduct, including mission planning, briefing, operations, debriefing, and maintenance. If schedule constraints are high, the control structure can be scoped down to include at least the items involved in one operating phase. Influences from outside the phase can still be included in the ensuing hazard analysis if an extended method like STPA-RC is used.

Developing the SDT control structure and choosing the appropriate test-framework strategy is as much a technical-planning as a safety-planning issue. Testers must ask themselves what information needs to be measured and communicated to support the test objectives. The methods, modes, and time scales of the control and assessment techniques required for testing shape the control structure. The visual diagram should itself distinctly reflect the aspects of the SDT that are legacy (or mimic legacy or field items), IBE, and part of the test framework. The visual format introduced in the previous chapter for the control structure assigns labels to the control actions (CA), feedback (FB), communications (CC), and indirect measures (IM) which can be quickly referenced in any discussion.[79]

Figure 4-12 shows the safety-control structure for the SDT of the wingman project. It focuses on the aspects of the operating phase and includes an additional simple control loop (top right of the figure) to include optional maintenance- phase considerations in the hazard analysis. The highest entity in this control structure is the air boss, a management representative for the test wing with the authority to dictate which actively airborne missions get priority of the airspace if the daily flying schedule becomes limited. The air boss also manages any conflicts between ATC and the test control room as they strive to keep air-traffic separated and ensure test mission success, respectively. ATC and the control room are both in radio contact with the test formation, and both receive radar position and velocity information from the formation. Physically, they both happen to be in the same building, but the control structure shows their functional relationships within

[79] I experimented with parenthetical cross-references to the control structure within the body of the STPA-based safety plan. However, the feedback I received from practitioners at Edwards what that this type of formatting detracted from reading the document. A more advanced technique like footnotes or hyperlinks might be warranted.

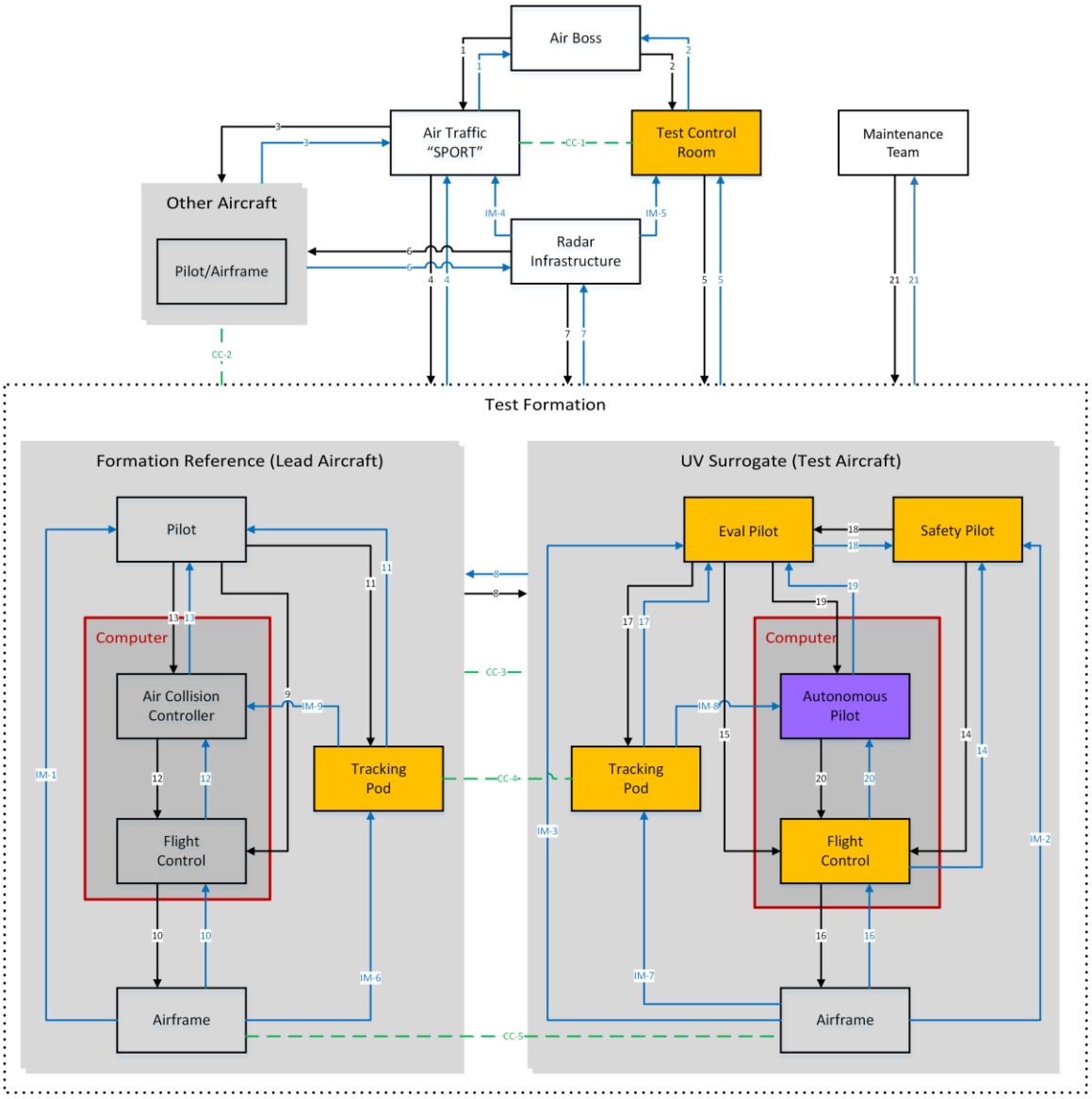



Figure 4-12. Autonomous Wingman Safety Control Structure

the SDT. Per my chosen format, Table 4-3 accompanies the wingman safety-control structure to provide the variable references.

When describing the SDT in the safety-plan documentation, the IBE, legacy aspects, and test framework are discussed one at a time. The safety-control structure is color coded to increase its use as a visual aid for this purpose. Purple entities are (or contribute to) the IBE. For the wingman project, the IBE is the autonomous-pilot algorithm.

Table 4-3. Autonomous Wingman Variable Reference

| Variable | Name | Variable | Name |
|---|---|---|---|
| **Control 1/2a:** | Priority Instructions | **Feedback 1/2a:** | Request for Priority |
| **Control 3/4a:** | Airspace Management | **Feedback 3/4a:** | Confirmation of Instructions |
| **Control 5a:** | Test point cadence | **Feedback 3/4b:** | Airspace Requests |
| **Control 5b:** | Troubleshooting help | **Feedback 3/4c:** | Position Reports |
| **Control 6/7a:** | Position/vel request | **Feedback 5a:** | Test Telemetry |
| **Control 8a:** | Test Cadence/Formation Calls | **Feedback 5b:** | Confirmations of instructions |
| **Control 8b:** | Contingency calls | **Feedback 5c:** | Troubleshooting questions |
| **Control 9a:** | Spatial Control Inputs | **Feedback 6/7a:** | Radar Return |
| **Control 10a:** | Control Surface Deflections | **Feedback 8a:** | Standard responses |
| **Control 11a:** | Pod Settings | **Feedback 8b:** | Contingency responses |
| **Control 11b:** | Formation Button Request | **Feedback 10a:** | Control Surface Positions |
| **Control 12a:** | Collision Manuever Request | **Feedback 10b:** | Aerodynamic State |
| **Control 12b:** | Collision Maneuver Type & Geometry | **Feedback 11a:** | Pod status |
| **Control 12c:** | Collision Maneuver Terminate | **Feedback 12a:** | Aerodynamic State |
| **Control 13a:** | Collision Settings (on/off/options) | **Feedback 13a:** | Collision maneuver indicator |
| **Control 14a:** | Spatial Control Inputs | **Feedback 13b:** | Collision system status |
| **Control 14b:** | Test mode engage/disengage and test pilot emer override | **Feedback 14a:** | FLCS Mode Status |
| **Control 15a:** | Spatial Control Inputs | **Feedback 16a:** | Control Surface Positions |
| **Control 15b:** | Manual disengage and test pilot emer override | **Feedback 16b:** | Aerodynamic State |
| **Control 16a:** | Control Surface Deflections | **Feedback 17a:** | Pod status |
| **Control 17a:** | Pod Settings | **Feedback 18a:** | Acceptance of test mode |
| **Control 18a:** | Permission/remission of test mode | **Feedback 18b:** | General Troubleshooting |
| **Control 18b:** | General Troubleshooting | **Feedback 19a:** | Autopilot status |
| **Control 19a:** | UV Autonomy Mode | **Feedback 19b:** | Collision maneuver indicator |
| **Control 19b:** | UV Receive Mode | **Feedback 19c:** | Collision System status |
| **Control 19c:** | Formation Button Request | **Feedback 20a:** | Aerodynamic State |
| **Control 19d:** | Turn Setting | **Feedback 21a:** | Inspections and Diagnostics |
| **Control 19e:** | Altitude Offset | **Feedback 21b:** | Diagnostics |
| **Control 19f:** | Collision Settings (on/off/options) | | |
| **Control 20a:** | Spatial Control Inputs | **Indirect Measure 1a:** | Motion |
| **Control 20b:** | Collision Manuever Request | **Indirect Measure 1b:** | Aerodynamic State |
| **Control 20c:** | Collision Maneuver Type & Geometry | **Indirect Measure 1c:** | TSPI and ranging |
| **Control 20d:** | Collision Maneuver Terminate | **Indirect Measure 2a:** | Motion |
| **Control 21a:** | Repairs | **Indirect Measure 2b:** | Aerodynamic State |
| **Control 21b:** | Software Configurations | **Indirect Measure 2c:** | TSPI and ranging |
| | | **Indirect Measure 3a:** | Motion |
| **Comm 1a:** | Maneuver/Airspace Justifications | **Indirect Measure 3b:** | Aerodynamic State |
| **Comm 2a:** | Sight of other vehicle in airspace | **Indirect Measure 3c:** | TSPI and ranging |
| **Comm 3a:** | Sight of other vehicle in formation | **Indirect Measure 4a:** | PV info for all players |
| **Comm 4a:** | TSPI of Other Ship | **Indirect Measure 5a:** | PV info for all players |
| **Comm 4b:** | Formation Request (from Lead) | **Indirect Measure 6a:** | TSPI of ownship |
| **Comm 5a:** | Radar/Transponder range signals | **Indirect Measure 7a:** | TSPI of ownship |
| | | **Indirect Measure 8a:** | TSPI of all players |
| | | **Indirect Measure 8b:** | Formation Request (from Lead) |
| | | **Indirect Measure 9a:** | TSPI of all players |

Orange entities in the SDT control structure constitute the test framework and would not be part of the fielded system.[80] The flight-control portion of the flight computer software is orange in Figure 4-12 because it is specially modified to operate in both a test mode (evaluator-pilot control) and override mode (safety-pilot control), as well as take commands from the autonomous pilot when the evaluator pilot engages it to do so. The two pilots on the surrogate UV aircraft are orange because they simply would not exist in the fielded concept of an autonomous wingman. Those pilots during test exist to manage the individual maneuvers being used to evaluate the algorithm, and to provide additional safety.[81] The datalink pods (called tracking pods in the figure) are modified for testing, and the test control room does not exist in the field.

The remainder of the entities in the figure are not in color, and they represent aspects of the system that resemble, duplicate, or emulate real-world fielded aspects. The lead aircraft, for example, is mostly a legacy asset except the special formation requests it can send through its pod to the wingman. Being able to visualize the test framework within the SDT provides an instantaneous appreciation for the need for test safety as a unique practice. Even when the design stage provides a control structure and STPA analysis for the product, testers must perform STPA Step 1 and 2 on the new control relationships that exist due to the updated safety-control structure, as well as revisit design-stage UCAs for new contexts that exist due to the updated safety-control structure (particularly the test framework).

Part-whole decomposition is present in the wingman-project control structure. The test formation is composed of the lead aircraft and test aircraft. Each aircraft is composed of pilot(s), a flight computer, datalink pod, and the airframe physics. Each flight computer is abstracted into two functional parts: the flight-control laws, and a higher-authority software loop that can give commands to them. Physically, however, a flight computer is itself just a bundle of hardware and software. Also of note is my decision to include multiple distinct communication channels between the two aircraft. CA-8 signifies that the lead aircraft has responsibility for the formation, even if the lead pilot cannot always see the wingman. CC-3 is the visual sight, when available, between the vehicles in the formation. CC-5 is the radar and transponder information the aircraft can use to range each other. Either aircraft's pilots can use CC-3 or CC-5 to update their mental models. CC-4 is the TSPI being shared between datalink pods.

Other aspects that would be in the control structure if the product testing required them would include, for example, entities like range controllers, RSOs, air and ground targets, and telemetry-communication equipment. As discussed earlier, the test enterprise should produce safety-control structures (at least in part) for its commonly available test-framework and organizational aspects, and practitioners are encouraged to perform STPA on these common aspects outside of any specific product evaluation.

---

[80] Orange items are either non-existent in the field, or they are not configured or modified in the field the way they are during DT.

[81] The safety pilot is highly experienced in the aircraft type, able to interpret (often non-designed) feedback well to detect imminent hazards.

Another part of the SDT discussion in the safety-plan documentation is a discussion of the system modes. Like the control structure, this knowledge is fundamentally created in the design stage and modified by testers. The wingman project is a good example. Due to the nature of the IBE, multiple modes of autonomy and coordination between and within the two aircraft existed to accomplish the evaluations and the test-point setups required for them. I had to read the traditional technical documentation several times to gain a basic comprehension of the modes for the IBE as well as each aircraft as modified for testing. I developed a new type of visual aid to help planners organize their understanding of the system modes and their relationships, called a *modes diagram*. The visual aid not only helps with designing the test activities, but can aid during the hazard analysis if STPA-RC is being used to examine controllers' process models.

Figure 4-13 shows the modes diagram for the wingman project. The diagram was built using the same software as the EIM (see Appendix B), although any visualization method could be used. The light-blue boxes show the commands of the evaluator pilot in the surrogate, the green box shows the commands of the safety pilot in the surrogate, and the gray boxes show the lead pilot's commands. Those command boxes include parenthetical cross-references to their respective control-action variables in the safety-control structure. The yellow boxes show the different possible selectable modes, while white boxes denote selectable behaviors. The white boxes would be optional for a modes diagram, but here they paint a fuller picture of the command authority available to the pilots in the system. The diagram could have been expanded to capture higher abstractions (e.g., ATC modes), but it was limited to the scope shown.

#### *4.3.3.4 STPA Test Safety Mitigations*

In the same manner that the wingman-project safety mitigations from the traditional safety plan were reproduced in Section 4.3.2.2, the safety mitigations from the STPA-based test-safety plan are reproduced in this section. A placeholder is included where a parenthetical reference might reside right after the first listed developing influence. As discussed earlier, every MP (developing influence, setting/configuration, operating procedure) is traceable to one or more UCAs and/or causal scenarios, and hence one or more hazards as well. This way of documenting traceability is difficult to implement with a text document, and the placeholder is only shown for the one MP and omitted for the rest.

Like the traditional safety mitigations, changes or paraphrasing are done where necessary to mask sensitive information about the project:

*System Notes and Restrictions*

With the surrogate test mode engaged, the UV surrogate is limited to [specific proportion] of the normal aircraft design limit load, and the maximum operating airspeed is [*removed*]. The autonomous-pilot evaluation is limited to this reduced envelope.

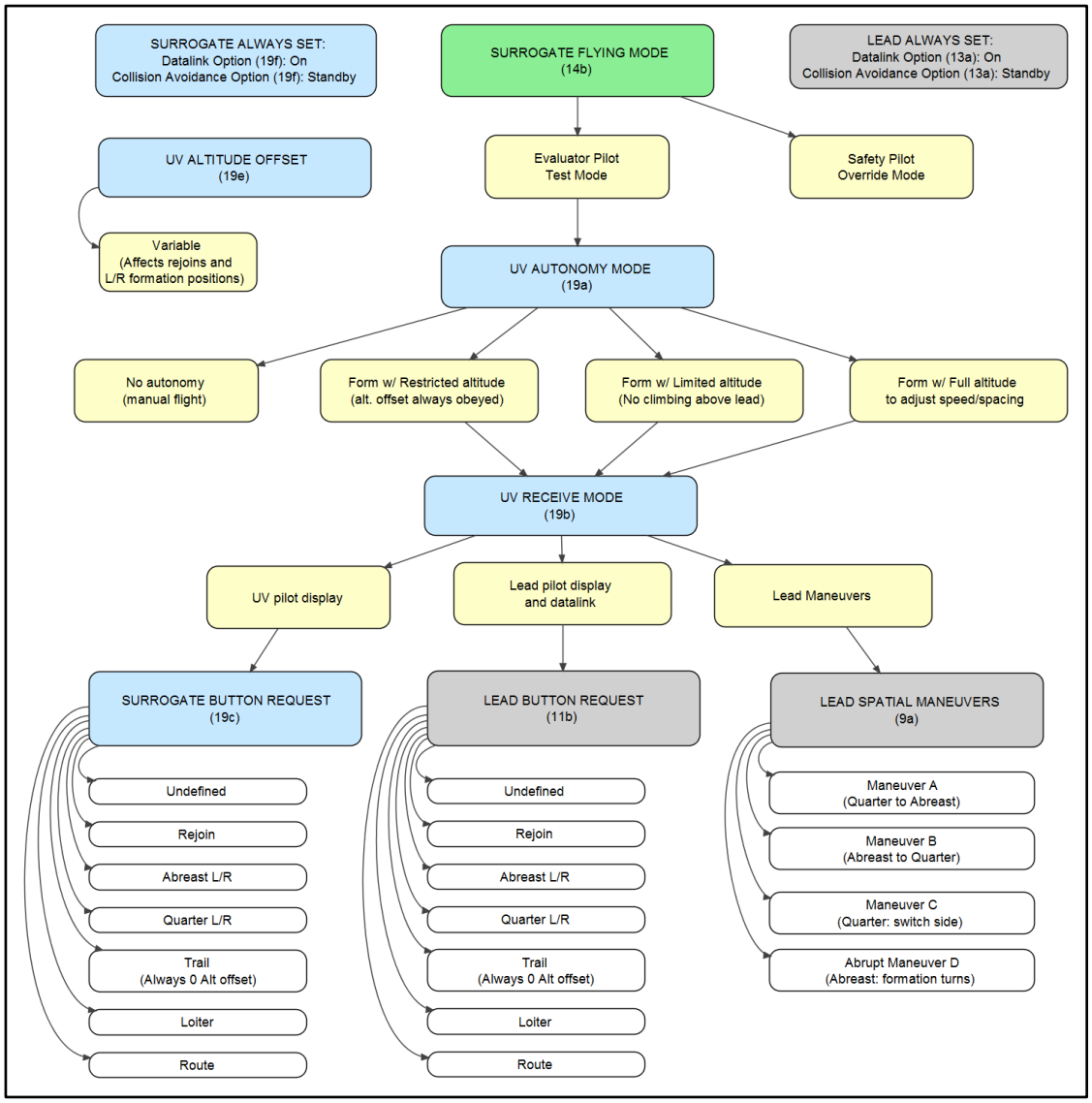



Figure 4-13. Autonomous Wingman System Modes

Wing datalink pods might be susceptible to datalink dropouts at [*removed*] degrees azimuth and [*removed*] degrees elevation off the wing if the pods are loaded on opposite sides between the two aircraft; less so if pods are on the same-side.[82]

There is a pod radiation hazard of [*removed*] feet for ground personnel when the pod has power.

[82] Pod limitations were discussed in detail in Section II of the safety plan and are revisited here as a restrictions; the limitation is also addressed below with a setting/configuration mitigation.

The autonomy algorithm is not programmed to perform formation turns in the direction of lead to wingman.

*Testing Restrictions*

Minimum separation distance between aircraft during all test points will be 500 feet.[83]

*Developing Influences*

*Test/Safety Planning*

1. Although the wingman aircraft is capable of a larger aerodynamic envelope, all test points will be planned between 10,000-20,000 feet MSL and [specific airspeed limits].

   *(List UCAs, list hazards)*[84]

2. All test points will be flown within the current modified aircraft guidance.
3. All test points will be planned well outside of the air-collision trip parameters.[85]
4. Safe lateral and vertical separation will be built into the test points to ensure minimum safe separation.
5. All autonomous route waypoints will be planned so that turn overshoots do not bring the UV surrogate near any airspace boundaries.[86]
6. Loiter points will be planned away from airspace boundaries.
7. Photo chase will only be executed on test points that have already been completed on a previous sortie and assessed to have no objectionable qualities, based on pilot comments and engineering analysis.

*Training and Qualifications*[87]

8. One pilot in the formation will be a current and qualified TPS instructor pilot.
9. The autonomous-pilot algorithm will be evaluated by all project test pilots in the simulators.
10. To qualify to pilot the lead aircraft, the pilots on the test team require checkout flights in the aircraft type. One formation checkout flight is required. These checkout flights will occur per the TPS local instructions.
11. Only a qualified [aircraft-contractor pilot with a high-level of experience] will serve as safety pilot in the wingman aircraft.

---

[83] This type of statement was an MP within the THA of the traditional plan. It is better documented here as a testing restriction, as it establishes the criteria by which mitigations that are more directive in nature (e.g., "if minimum separation is violated, do X") can be enforced.
[84] Placeholder to demonstrate the optional use of parenthetical traceability.
[85] This has now been made an explicit requirement instead of a passing statement elsewhere in the document.
[86] Both this and the subsequent operating procedure mitigate hazard (4).
[87] Many of these have been split off and itemized from the paragraph format of the traditional version

12. To qualify to pilot a photo/chase aircraft, a pilot will have to complete the photo/chase upgrade syllabus. The photo/chase pilot will not be a TPS student.
13. The control room will be occupied by project members not flying and designated aircraft engineering contractors; anyone else must be familiar with this safety plan and attended the mission briefing.

*Flight and Test Manuals*

14. The current modified aircraft guidance will be available to the test team.

*Settings and Configurations*

*Test Card Requirements*

1. Each test point's autonomy mode, receive mode, intended maneuver, and altitude offset shall be included on each test card.

*Briefing Requirements*[88]

2. Formation roles, responsibilities, and blind procedures will be in accordance with AFI II-2FTV3 and briefed before every sortie (emphasis on blind communications and deconfliction)
3. All maneuvers, areas of concern, responsibilities and terminology will be covered in the flight briefing.
4. Flight briefing will discuss use of the aircraft transponder and radar for distance finding and ranging (ref CC-5 in control structure).
5. Any missions with photo chase will discuss photo chase positioning during the mission briefing.[89]
6. Maintenance personnel will be briefed before each flight by the aircrew, adhere to pod safety distance, and follow all applicable pod handling guidance.

*Instrumentation and Item Configurations*

7. The latitude, longitude, altitude, and airspeed for each autonomous route and loiter routine will be entered via a data file loaded into the wingman flight computer and confirmed before each flight (the parameters cannot be adjusted real time).[90]
8. Datalink pods shall be loaded on same-side wing on both aircraft.[91]
9. The air-collision algorithm will be turned off (standby) and confirmed so in both aircraft prior to taxi.

[88] Many of these have also been split off and itemized from the paragraph format of the traditional version.
[89] Photo-chase briefing was not an explicit requirement in the traditional version.
[90] This has now been made an explicit requirement instead of a passing statement elsewhere in the document.
[91] This has also now been made an explicit requirement instead of a passing statement elsewhere in the document.

10. The following will be checked for functionality and verified operational prior to taxi:
    - Cockpit Intercom
    - Radar and transponder
    - Radios (primary and backup frequencies)
    - Pod communication signals
    - Control/displays of wingman flight-control mode

*Special O&M Considerations*

(None)

*Special ORM / Physiological Considerations*

(None)

*Operating Procedures*

1. All testing will be conducted in day VMC. All aircraft must be able to maintain 2,000 ft. vertical and l NM horizontal cloud clearance with 5 nautical mile visibility and a discernible horizon. (ref CC-3/CC-2 in control structure)
2. Both aircraft will adhere to AFI 11-2FTV3 for formation roles and responsibilities.
3. Wingman evaluator pilot will use maneuvering as necessary during test-point setup to remain within the test-mode envelope prior to and during autonomous-pilot initiation.
4. Prior to each test maneuver, the air-collision algorithm will be verified to be in standby on both aircraft.[92]
5. At no point will the autonomous pilot or air-collision algorithm be relied upon to maintain safe aircraft separation.
6. The datalink option will be turned on and confirmed in both aircraft prior to first test point.
7. Altimeter calibration checks between aircraft will be performed on departure and at altitude prior to the first test point. Checks will be performed with the surrogate UV in manual control, 10 ft. spacing and stacked level. Both aircraft will set to 29.92. If the lead differs from the wingman by more than 50 ft., lead will adjust its altimeter setting until its altitude matches the wingman.
8. At no point will the aircraft be closer than 10 feet apart.
9. All maneuvers shall be executed with sufficient lateral and/or vertical separation to provide room for either aircraft's pilot to safely maneuver and avoid the other aircraft by the minimum separation distance.

[92] This statement was separated out from a similar statement in the settings/configurations.

10. Lead will coordinate with ATC to ensure the airspace between the wingman loiter point and lead's position is clear of traffic at all altitudes before a rejoin is commanded.[93]
11. Lead will ensure formation is always beyond 5 nautical miles from any airspace boundary on either side.
12. Aircrew between the two aircraft will confirm autonomy mode, formation position, and test point via challenge-response prior to executing each point.
13. When autonomy is in lead push-button request mode, lead will confirm with wingman aircrew that request was received by the autonomous pilot.[94]
14. Aircrew within the wingman will confirm autonomy mode, receive mode, intended maneuver, and altitude offset via challenge-response prior to executing each test point.
15. The wingman evaluator pilot will read out altitude-offset setting to the safety pilot for every new entry to confirm fourth zero is entered and accepted.[95]
16. At least one aircraft must always be visual of the other during formation activities and within 2 nautical miles in rejoin.
17. Wingman pilots will follow standard communication and CRM for transfer of aircraft command and test-mode enabling/disabling.
18. Photo chase will primarily base its positioning off of the lead aircraft. The intent is for photo chase to remain on the outside and/or aft of the formation (keeping the lead aircraft in the middle of the formation). At no point shall photo chase close within 100 ft. of either aircraft.
19. At no point will lead attempt an aggressive turn toward the autonomous wingman when flying wing abreast.[96]
20. The wingman will perform all formation points offset 200 ft. below lead's altitude to provide room for either pilot to safely maneuver and avoid the other aircraft if necessary.
21. The wingman will perform all rejoins offset 500 ft. below lead's altitude to provide room for either pilot to safely maneuver and avoid the other aircraft by at least 500 ft.

---

[93] Both this and the subsequent operating procedure mitigate hazard (4).

[94] I determined from studying the design and speaking to testers that the lead aircraft does not provide display feedback to its pilot that the wingman has received push-button requests.

[95] I determined from studying the design and speaking to testers that the control panel requires an extra number's place in the offset entry, and the difference between having entered three or four digits might not be noticeable in the appearance of the confirmation screen.

[96] This is due to a limitation in the autonomy algorithm; it was not configured to adhere to the TTPs for situations in which the formation is wing-abreast and lead turns toward the wingman. Turns away from the wingman were possible.

22. To ensure the VISTA does not exceed the 500 ft. altitude buffer, 600 ft. will be entered via the flight-control "altitude offset" entry.

*Hazard Corrective Actions (CTA)*

*H1: Aircraft violates minimum separation distance to other flying objects*

1. A test point will be terminated if the photo chase loses sight of either aircraft.
2. If the wingman in formation goes "blind", the test point will be terminated, and standard communications shall be used until regaining a "visual" on target and establishing positive deconfliction. The wingman in this case will be flown by the safety pilot with the test mode disengaged.
3. In rejoin maneuvers, if neither aircraft gain sight of each other by 2 nautical miles, the point will be terminated, and the safety pilot in the UV surrogate will disengage the test mode, call blind, maneuver away from the target's last known position, maintain current altitude, and begin coordinating a rejoin via radio communication with the lead aircraft.
4. If the respective required altitude offset is violated on rejoins and formation points, or should the wingman aircrew perceive a threat to minimum separation, the safety pilot will disengage the test mode immediately and maneuver the aircraft appropriately.

*H2: Aircraft violates terrain closure limits*

1. Test point will be terminated.
2. Safety pilot will disengage the test mode immediately and maneuver the aircraft appropriately

*H3: Aircraft departs aerodynamically stable flight*

1. Safety pilot will disengage the test mode if not already deactivated and perform recovery procedures.

*H4: Aircraft exits allowable testing area*

1. Test point will be terminated and lead will coordinate with ATC to reestablish vector toward allowable boundaries.

*Mishap Recovery Actions (RA)*

*Accidents A1–A4*

1. Surrogate test mode will be disengaged
2. If the wingman safety pilot is incapacitated, the wingman aircraft will be put into emergency mode.[97]

[97] This is a flight-control setting available to the evaluator pilot that allows the evaluator pilot to fly the aircraft with the full aerodynamic envelope and control authority normally allotted to the safety pilot.

3. The formation aircraft will follow all relevant technical-manual emergency procedures, performing flight controllability checks (if applicable) to determine any controllability and handling qualities degradation or limits prior to returning to the airfield.
4. If required to execute a controllability check, aircraft will request an in-flight visual inspection if practical. The aircraft conducting the visual inspection shall not have been involved in the collision.
5. The control room (if used) and the TPS superintendent will coordinate recovery assets with the air boss and base emergency services per local policy and guidelines.

*Accident A5*

1. Aircrew will coordinate directly with airspace controllers to correct their heading.

*Accident A6*

1. If no other accidents have occurred, test team will discuss further test point attempts if data collection is still possible.

### 4.3.4 Comparison of Methods and Mitigations

Both the STAMP and the DOD safety mindset evolved from the system-safety movement of the twentieth century. The second and third system-safety processes of MIL-STD-882 prescribe the assessment of safety risk and identification of risk mitigation procedures, respectively, once hazards are identified [11].[98] When it comes to the general philosophy for reducing the safety risk in a system, the STAMP guidance maintains focus on the hazards; it implements reduction methods in this order of precedence [5]:

STAMP Risk Reduction

1) Design system to eliminate hazard
2) Reduce hazard likelihood
3) Control hazard exposure
4) Lessen damage severity

The first three methods apply generally to system hazards, while the fourth addresses mishaps. Table 4-4 shows the risk reduction methods in the first column; some of Leveson's suggested approaches and considerations for implementing each method are in the second column [5]. AFTC has its own order of precedence for risk reduction, based on AF policy [29]. That list of methods is almost as short as Leveson's, but it also includes some suggested implementation approaches [24]:

AFTC Risk Reduction

1) Design system (or test) to eliminate the hazard

[98] Refer to Section 4.2.2.1 for a review.

Table 4-4. Comparison of Risk Reduction Approaches [5], [24]

| Systems Risk Reduction | STAMP Approaches (Leveson) | Air Force Approaches (AFTCI 91-203) |
|---|---|---|
| Design to Eliminate Hazard | ➢ Substitute safer materials<br>➢ Simplify (minimize) parts, modes, interfaces<br>➢ Decouple interdependent components<br>➢ Reduce human ability to initiate hazard | AF 1 - Design system or test to eliminate hazard* |
| Reduce Hazard Likelihood | ➢ Passive – failsafe (restrictive, relies on physics, not always feasible)<br>➢ Active – detect and correct (may be complex and requires diagnostics)<br>o Failure warnings and indicators<br>o Decision aids<br>o Trained and practiced procedures<br>➢ Lock-out/Lock-in Barriers: physical or software limiters<br>➢ Interlock: enforces correct sequence or inhibits bad combos<br>➢ Structural factors of safety<br>➢ Redundancy, spares, voting | -----<br>AF 2: Change test methodology<br>-----<br>AF 4: Caution and warning devices<br>AF 5: Procedures and training |
| Control Hazard Exposure | ➢ Limit exposure time (e.g., master arm switch)<br>➢ Isolate/contain hazard from the environment<br>➢ Fail-safe states which are easy to get into<br>o Panic button or kill switch<br>o Watchdog timer<br>o Passive devices (e.g., dead-man switch, door magnets)<br>o Intended failure paths (e.g., crack before burst, engine falls off wing)<br>o Graceful degradation, multiple failure states<br>➢ Salient alarms and status messages (beware alarm fatigue) | -----<br>AF 2: Change test methodology<br>AF 3: Engineering and safety devices<br>AF 4: Caution and warning devices<br>AF 5: Procedures and training |
| Lessen Damage Severity | ➢ Defense in Depth (e.g., facility layered walls and vessels)<br>➢ Escape routes (e.g., ejection, RSO destruct command)<br>➢ Damage limiters (e.g., blowout panels, sheer pins)<br>➢ Emergency training (sims or drills)<br>o Threat recognition, Decision points | -----<br>AF 2: Change test methodology<br>AF 3: Engineering and safety devices<br>-----<br>AF 5: Procedures and training |

*MIL-STD-882E: "No amount of doctrine, training, warning, caution, or personal protective equipment can [eliminate] a mishap probability" (p. 11)

2) Change the test methodology to reduce mishap probability and/or severity
3) Incorporate engineering and/or safety devices (e.g., parachute, redundant power)
4) Provide caution and/or warning devices to detect and unsafe condition or trend
5) Develop procedures and training when the above are impractical

In the AFTC list, only the first two items are general risk-reduction methods. The first item matches the first STAMP item (design to eliminate the hazard).[99] The second AFTC item—change the test methodology—attempts to capture everything else on Leveson's list (reduce hazard, control hazard, and lessen severity). The last three AF items are merely

[99] Both Leveson and the AF emphasize that designing to eliminate the hazard means altering the system so that the hazard is physically not possible (see discussion in the previous chapter about Hajdukiewicz et al.'s physical constraints [170]).

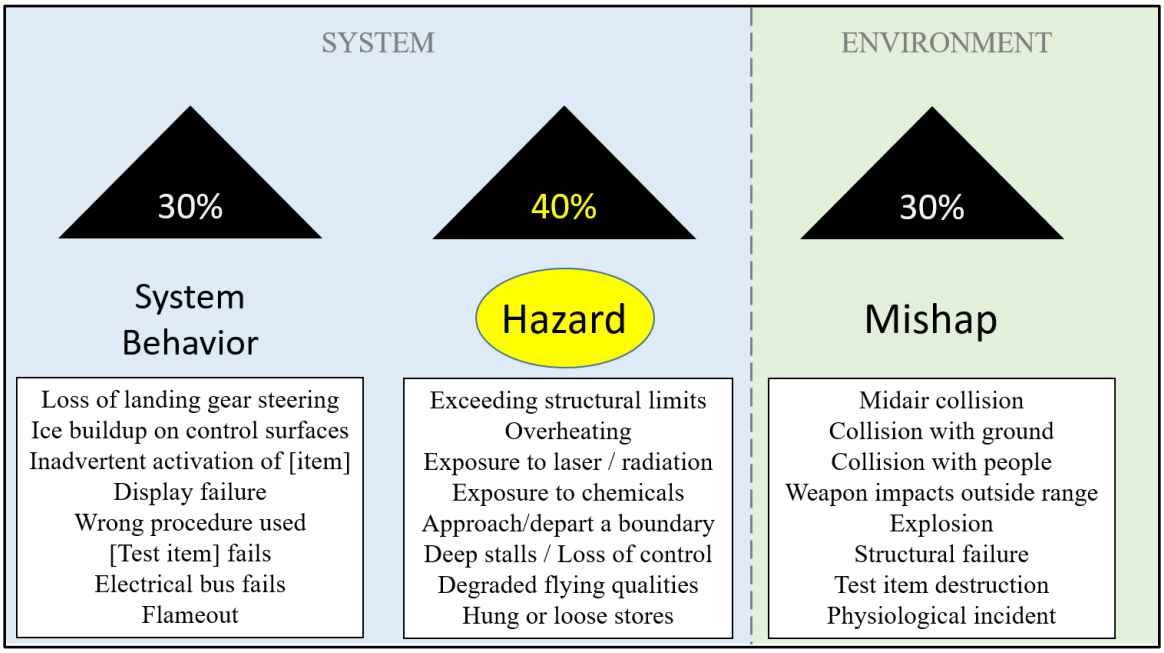



Figure 4-14. 412th Test Wing THA Review using STAMP Criteria

engineering and operating approaches that are intended to implement the second method (change the methodology). Although the AF recommends the order of precedence presented above, Table 4-4 attempts to put the AF items into a third column to show that they can apply in many places with respect to Leveson's order of precedence.

The concession in AFTC that allows for some mishaps to be identified as hazards (see Section 4.3.2) is a byproduct of the AF event-chain mentality for determining risk, and it tends to conflate hazards and mishaps. Figure 4-14 shows the findings of a review of THA worksheets I performed at Edwards. I thoroughly reviewed the several-hundred archived THAs in the wing safety office, and based on their titles and descriptions I determined whether they met the criteria of a hazard as defined in STAMP. I gave the benefit of the doubt when a test hazard appeared to describe a system-level hazard. Even then, only 40 percent of the test hazards met the criteria. The remainder of the test hazards were not only mishaps; half of them were actually very specific UCAs or causal scenarios (by STAMP criteria).[100] The figure gives some examples of THA titles in each of the categories of system behavior (UCAs/causal scenarios), hazard, and mishap.[101]

The risk-based mentality and the ambiguity it creates in documenting test hazards fosters a piecewise process for applying engineering and safety devices, cautions and warnings, and procedures and training. The proposed STPA-based test-safety planning format simply follows Leveson's risk-reduction methods after accidents and hazards have been identified. MPs reduce the hazard likelihood, CTAs control hazard exposure, and RAs lessen damage severity. Engineering and safety devices, cautions and warnings, and

[100] This also means that those test hazards were *very* specific to the design or testing.
[101] Sensitive data have been removed, and the total number of THAs in the archive is not shown.

procedures and training may be applied in a complementary fashion—as needed with no priority of one approach over the other—with a systems view aided by the functional control structure of the SDT. Eliminating a hazard is not possible without modifying the control structure.

Ideally, the STPA-based approach to product safety begins in the design stage. As such, testers do not have as much latitude as designers to affect product-specific portions of the safety-control structure.[102] However, testers do have major bearing over the test framework, and they can shape the SDT in that way. Risk reduction becomes increasingly limited the further a product is into a detailed design and manufacture. The same is true for how far the planning of a test project has progressed. It is easier to add, remove, or rearrange functional components of the SDT (in order to possibly eliminate hazards) early on during technical test planning. Later the options are reduced, and mitigations might not be as well thought-out or efficient (e.g., add-on displays or extra training).

For cases in which the purpose of the test is to reduce the risk of a new technology that has knowledge gaps or poorly substantiated design assumptions,[103] the technical plan should be very clear as to what hazardous scenarios might be likely (or more likely) to occur as the test sorties build up the performance and maneuvering envelope of the IBE. A straightforward technical plan that can already begin to reference a safety-control structure and discuss these scenarios is important to have when performance objectives *are* safety objectives. "Types of Tests" in Section III of the STPA-based safety plan, where UCAs and causal scenarios are documented (or at least referenced) is guided by the comprehensive technical planning of the evaluations and maneuvers to be used during testing.

There are also cases where the purpose of the test is itself to assess a safety mitigation.[104] For these cases, testers may even intentionally introduce a hazard into the SDT. This is certainly not something that would happen in the field stage. However, the systems view can handle this. The hazard, which is being intentionally created (e.g., loss of aerodynamic stability), is documented as such. While the safety mitigation being evaluated would normally be identified as a CTA, it is instead—for this specific type of testing—the IBE. Other types of CTAs that are unrelated to the IBE must be identified as mitigations in the safety planning. MPs do not apply (except to prevent the hazard or reduce its likelihood for portions of the test mission when it is *not* wanted).

It should be re-emphasized that even though both the traditional and STPA-based planning approaches produce safety mitigations, they treat the identification of hazards fundamentally differently. Test hazards do not exist in the STPA-based approach, because it does not typically consider hazards to be specific to the design or testing of the product. Rather, it acknowledges that all hazards are at the system level. This does not result in a

---

[102] Many of the risk-reduction approaches in the second column of Table 4-4 are meant for designers and are not available to testers.

[103] The wingman project is an example of testing in which several of the technical performance assessments were accomplishing to system-safety risk reduction.

[104] For example, an aircraft may be intentionally put into an unstable aerodynamic condition to evaluate an emergency flight-control command that rights the aircraft.

Table 4-5. Comparison of Hazard Analysis Methods [238]

| | Traditional | STPA |
|---|---|---|
| **Analysis Philosophy** | Success oriented<br>*(i.e., it assumes nominal case then tries to predict probability of deviation)*<br><br>Provides set of contingencies for off-nominal behavior | Assumes worst-case scenario<br>*(i.e., it starts with accident, then hazards, then causal factors, and assumes that any of the causal factors can happen)* |
| | Emphasis on preventing or reducing failures | Emphasis on enforcing constraints on system (and thus component) behavior |
| | Assumes most failure modes are independent | Accounts for sub-system interactions and how these influence safety-related behavior |
| **Causal Factors** | Considers only hardware failures, or treats operators and software as if they are hardware<br><br>*(e.g. leaves on a fault tree with assigned probabilities of failure)* | Assumes that software does not "fail" but can still be hazardous due to *flawed requirements or unsafe interactions* with rest of system<br><br>Human operators perform within the context of a larger system design and, like software, do not necessarily "fail" but can make unsafe decisions |

loss of the information traditionally contained in THAs. Instead of mitigations being imagined as failure preventions along a foreseeable chain of events, they are treated as constraints on behavior within a control structure. Instead of being organized in the document by the specific test hazards they target, they are consolidated into non-repeating statements that are arranged by the type of influence (developing influence or setting/configuration) or real-time control (operating procedure/corrective action) they are, and traceability of each mitigation to multiple system hazards is allowed.

Fleming et al. presented a very straightforward tabular comparison between probability-based and systems-based hazard analyses in their report of the In-Trail Procedure [238]. Part of the information is reproduced in Table 4-5 to emphasize the main differences between the hazard analyses underlying the traditional and STPA-based test-safety planning methods. A more detailed table can also be found in Leveson et al. [184].

#### *4.3.4.1 Flight Test Project Safety Plans*

In the previous sections, the document formats for both the traditional and proposed STPA-based safety-plan methods as well as some of their outputs (including safety mitigations) for the wingman project were presented. To reiterate, the TPS students wrote the traditional safety plan as well as the technical plan; they had access to simulators and discussions with designers to reinforce their understanding of the SDT. I did my best to become familiar with the wingman system by reading the technical plan and Section I of the traditional safety plan and asking the students for clarifications, but my knowledge was not as detailed. The language in the STPA-based plan is intentionally written to be very similar to that in a traditional plan, and I used several older test-wing planning documents for reference. This had the benefit of allowing the STPA-based document to assume a familiar tone for testers, despite it being different in structure. Comparisons between the two methods along several important attributes are shown next.

*Planning Commitment*

The TPS students reported having spent approximately ten hours to construct the traditional safety plan. I spent approximately 60 hours to construct the STPA-based safety plan. This number merits discussion. Five hours were spent constructing the safety-control structure. This took time because I had to become familiar with the project as I worked on the safety plan. I built an initial safety control structure and revised it several times as my knowledge of the SDT increased. Ideally, a team of discipline engineers and test operators would work on the project and have a much fuller understanding of the system before safety planning begins.[105] Additionally, I used standard office software to build the control structure, a tedious and inefficient format for long-term use. A better estimate for the time to construct a control structure for a project with the scale of the wingman testing and with a team of knowledgeable planners would be two hours, based on the time it took me to whiteboard diagrams and confirm assumptions with test-project members.

Forty hours were spent identifying and tabulating UCAs. Tables were built (per the format in Chapter 2) for this task and constructed manually in a spreadsheet. This method is not sustainable for long-term use, and a more formal database technique that can keep findings organized consistently is recommended. However, an improvement in the tools for building tables would only have reduced the time spent to perhaps thirty hours at the least. The identification of UCAs (and causal scenarios) is the heart of the STPA-based approach, guided by the hazards identified earlier. It is a substantial hazard analysis that creates more documented products than the traditional planning method.

Fifteen hours were spent writing the actual document. This would normally take less time as well, especially once templates are available within the enterprise. The proposed STPA-document structure was developed while drafting the notional wingman document. I spent a careful amount of time rewriting the project description (Section I of the traditional document and Section II of the STPA-based document) as I learned more about the wingman system. This involved organizing many small discussions scattered within the traditional document to follow the logic of the new structure by discussing test objectives, the system description (referencing the control structure), its limitations, and expected performance. Ideally, writing the report would take half the time. This puts the estimate of the total time that would have been required to perform STPA-based safety planning on the wingman project, under ideal conditions, at approximately 40 hours. That is four times that of the traditional method, and three quarters of it is spent identifying UCAs and causal scenarios.

The traditional document was 23 pages in length, while the STPA-based document was 32 pages (40 percent increase). The STPA document did not reproduce the full list of UCAs and causal scenarios in Section II nor did it include traceability statements from MPs to UCAs in Section III; it only presented small examples of each. Even so, 60 percent of the words printed in the STPA-based document were new language—not reused from

[105] Test planners would also ideally be provided the designer's control structure to serve as a basis for constructing the SDT control structure.

Section I of the traditional plan or any other older documents I referenced as a starting point.[106] Page length is not as much an issue in modern documentation as organization and intelligibility. The STPA-based method involved performing a substantial hazard analysis that resulted in tables that would be best documented in a separate database, away from the planning-document (unless printed in an appendix to the document).

*Results*

The traditional method identified one test hazard (midair collision during formation maneuvers). This hazard was determined during the THA process during safety planning. From a systems perspective, it is actually a mishap (collision), qualified by a causal scenario (formation maneuvers). The STPA-based safety planning used an *a priori* identification of accidents and hazards. In this top-down approach, I chose four hazards before beginning the analysis. They were (in short): violation of minimum separation between aircraft, violation of terrain closure limits, departure from aerodynamic stability, and exiting the testing area. The hazards would have notionally been provided and/or approved by the 412TW safety office had the project planning been real. Other hazards would exist in the safety-office repository, such as overheating, exceeding structural limits, exposure to chemicals, etc. However, hazards chosen to guide the analysis were based on initial discretion with regards to the general abilities of the system and the basic types of procedures being used in its operation. A cautionary note is appropriate here: the hazard analysis designers provide to testers would include the hazards the designers envision the system potentially having in the field. Testers should always consider the assumptions of the products and documentation they receive to include the appropriateness of the hazards the systems can encounter.

The traditional planning resulted in five GMPs and one THA. If the constituent parts of the THA are considered individually, a comparison of the analogous aspects of the plans can be made. This is shown in Table 4-6. The order of rows is based on the logical flow of the systems approach. Accidents and hazards were identified, followed by the hazard analysis to identify UCAs and causal scenarios, followed by mitigations. The three causes for the test hazard in the traditional approach were determined via some form of root cause analysis (likely brainstorming), while the large number of UCAs of the STPA were determined via systematic analysis of the safety-control structure. The MPs in the traditional format address specific causes of the test hazard or the testing in general. Every MP in the STPA-based format can be tied to one or more specific UCAs/causal scenarios, which are in turn tied to the safety-control structure and the system hazards.

Of the 392 UCAs identified, some were related to normal field procedures, but approximately 250 were related to the IBE and/or test framework. There could have well been more UCAs I did not identify due to the nature of the design of the wingman algorithm; this emphasizes the importance of communication between the design and test stage for a good understanding of a system. Conversely, some of the UCAs I identified could have been consolidated using more formal methods of analysis [189]. The important

[106] 4,940 of 8,210 words were mine, spread throughout Sections all three sections of the STPA document.

Table 4-6. Comparison of Safety Plan Format Results

| Traditional | STPA |
|---|---|
| 2 Effects | 6 Accidents |
| 1 Test Hazard (actually a mishap) | 4 System Hazards |
| 3 Causes | 392 Unsafe Control Actions |
| 13 Minimizing Procedures<br>*- 8 THA minimizing procedures*<br>*- 5 general minimizing procedures* | 46 Minimizing Procedures<br>*- 14 developing influences*<br>*- 10 settings/configurations*<br>*- 22 operating procedures* |
| *Nothing identified to control hazard exposure (test hazard was a mishap)* | 8 Corrective Actions |
| 1 Accident-Corrective Action | 7 Recovery Actions |

thing to note is the difference in the order of magnitude (two) between the number of causes in the traditional plan and the number of UCAs in the STPA-based plan.

Out of the 46 MPs in the STPA-based plan, 28 were exclusive to the STPA-based plan (60 percent).[107] There were also two new system restrictions that were not discussed in the safety-implementation section of the traditional plan. One was that the datalink pods would experience certain dropouts if not loaded onto the same-side wing on both aircraft. A corresponding setting/configuration was included for maintenance personnel to load the pods on the same-aide wings before every sortie. The other newly documented restriction was that the autonomy algorithm was not programmed to perform formation turns in the direction of lead to wingman. A corresponding operating procedure was included prohibiting those types of formation turns.

The footnotes in Section 4.3.3.4 provide more information about many of the listed STPA safety mitigations. Two of the more unique findings using the STPA-based method involved human controllers, aided by STPA-RC. I determined that the data-entry interface for the evaluator pilot in the surrogate aircraft to input the autonomous pilot's altitude offset required an extra zero (tenths place). This might introduce problems with affordance; it could be easy to forget to enter an additional digit because similar data-entry interfaces on other aircraft do not have this requirement. Forgetting the last digit would result in a safety buffer one tenth the size as expected. An operating procedure was added for the evaluator pilot to read aloud the entry with the extra digit as an enforcing action (the safety pilot had the confirmation display repeated on his console). I also determined that when the lead was

[107] Of these, there were nine developing influences, seven settings/configurations, and twelve operating procedures.

providing formation requests to the wingman via push-button actions through the datalink, there was no salient feedback to the lead pilot that the signal had been received. The hard-affordance feedback (the feel of the button being pushed[108]) could disagree with real events if the wingman did not receive or follow the request. An operating procedure was added for the aircrew in the surrogate aircraft to confirm via radio that the autonomy state changed on their own status displays.

There was only one safety mitigation in the traditional safety plan that was not documented in the STPA-based plan. That was a directive for all testing to be done at a minimum of 5,000 feet AGL. I provided altitude limits in the STPA-based plan as well, but only in distance above sea level. The traditional planners decided to use an AGL reference frame because some of the routes chosen for the testing were near mountainous terrain. This is a valid format for directing altitude limits, and it should be used for those circumstances. My intention with altitude limits was still to ensure terrain closure limits were not broken, but it was also to ensure the surrogate aircraft remained in a predictable aerodynamic envelope. In STPA, both terrain-closure and aerodynamic-stability hazards were acknowledged even though the traditional plan did not directly address them.

The one CTA documented in the traditional plan was in reference to its test hazard. It was a simple directive to "execute appropriate emergency procedures" if the test hazard (which was actually a mishap) were to occur. The STPA-based method produced eight detailed CTAs that addressed specific hazards. Four of them were similar to language that existed in various different sections of the traditional plan to address midair-collision prevention. The other four were new as they addressed terrain closure, aerodynamic stability, and exiting of the test area, all hazards not addressed directly in the traditional plan. Additionally, seven RAs were prescribed in the STPA-based plan to address its specific mishaps. The language was more directive and detailed, emphasizing the specific procedures that must be followed by different controllers in the hierarchy.

*Format*

The difference in philosophy and general document structure between the two types of safety plans has been discussed in previous sections. Although the constituent parts of the traditional THA were compared in the previous table to STPA mitigations, in reality a THA worksheet is meant to be an inclusive item for the purposes of the single hazard addressed. The wingman project had only one test hazard, but many larger test projects at Edwards have multiple THAs. Each one lists causes, minimizing procedures, and corrective actions affiliated with only the hazard in its title. These procedures and directives are often repeated across multiple THAs for the same project.

The MPs in the STPA-based plan are organized by developing influences, settings/configurations, and operating procedures. Each one is distinct and they do not repeat, and the MPs each trace to one or multiple hazards. This traceability is natural due to the top-down approach of STPA. Presenting that traceability in a written document, however, is not as straightforward. The traditional format uses parenthetical entries next to

[108] Or the change in lead's display that the command was sent.

each minimizing procedure in a THA to reference the test-hazard cause(s) it prevents. Using a similar formatting approach to trace the MPs in the STPA-based plan to UCAs would make the document unreadable. More tractable methods to reference the hazard analysis within the planning document could vary from footnotes to electronic hyperlinks—assuming the UCAs are kept organized in an accessible database external to the document.

The STPA-based safety plan culminates in distinct, non-repeating MPs, and an advantage there is that each MP can be addressed if it is deemed to overly-restrict test conduct.[109] By knowing exactly how many (and which) UCAs a questionable MP relates to, planners can discuss adjustments or other alternatives. Changes can then be made (and briefed to decision-makers) with a clear picture of its impact in the safety-control structure. During expert safety review meetings, the control structure may be used as a visual planning tool, and its labeled and numbered parts are easy to reference and use for discussion.

What is lost, however, in using the systems format for listing MPs is the ability to qualitatively highlight specific test scenarios. THA worksheets—although they are developed with much less rigor—group mitigations by these scenarios. Even though the systems approach does not support individual chain-of-event mitigations, the traditional format is able to provide a few salient test scenarios for experts to review and for aircrew and control-room personnel to brief before a mission and recall during operations. Currently, if a test hazard occurs during a test mission, the test project is suspended and a new safety review is triggered [24]. Using the STPA-based method would allow for this. Instead, if a *hazard* occurs during a test mission (when not done so intentionally), a new safety review could be triggered.

Currently, the review and approval of safety plans in the 412TW is as much an art as a science. Reviewers inject common sense from experience, and even with an STPA-based format they would still be encouraged to do so. However, the STPA-based format gives planners a deeper structure on which to draft the plan being reviewed, and it forces them to be more methodical in the hazard analysis. The documentation of results is logical and traceable. The new "overview of findings" was put into Section I of the proposed STPA-based format to give stakeholders a front-page summary of the hazard analysis in terms of the total number of hazards, UCAs, MPs, etc. This, along with the inclusion of remarks and considerations that planners deem useful to decision-makers, provides a concise summary that frames the larger document.

Both the traditional and STPA-based planning documents are prone to repeat substantial portions of the technical plan. The introduction of a system-theoretic format for describing the SDT was made in Section II of the STPA-based document. Ideally, that information might be just as well (or better) suited to be included in the technical plan. The functional-control structure and modes diagrams, for example, are useful and visual methods for guiding technical discussions as well as safety discussions. If a systems

[109] Often times (but not always) the test objectives are purely technical and not safety related. Test-safety planning is meant to determine the risk of test conduct performed to achieve the evaluation objectives.

approach provides a logical, top-down method of organizing information in a safety plan, it could potentially be used to improve how a technical plan is organized. Future thought should be given to how a technical and safety plan interface and overlap in information, as they are both parts of the larger test plan package.

The "predicted/expected results" part of the STPA-based safety plan (Section II) was the left the same as in the traditional safety plan (Section I).[110] It presented an optimistic view that the wingman algorithm would perform everything as designed with no surprises. Through fact-finding interviews at Edwards, practitioners told me that most test plans take this very positive stance when discussing predicted results. This mindset does not do well to foster a worst-case approach to the analysis. It also puts more onus on expert reviewers to bring up their concerns from experience (see footnotes of the test-safety discussion in Section 4.2.2.2 for a true-life example). Instead, planners should question the planning from the beginning. They should not be optimistic about system behavior, and they should perform a rigorous hazard analysis they can present to reviewers.

One aspect of the traditional planning method that is not present in the proposed STPA-based approach is an approach for the estimation of project risk level. In the traditional approach, a recommended overall risk is presented to reviewers, who can accept or modify it and before forwarding to approvers. The risk is determined by looking at each THA in the document and its estimated risk level while considering the perceived complexity of the test, knowledge gaps for the system, and personal experience [24]. The STPA-based test-safety plan does not recommend a risk matrix entry because STPA does not itself by nature calculate or estimate probabilities for mishaps and deviations from expected behavior (see Table 4-5).

This thesis does not recommend any one method of proposing an overall risk level for a test project. Leveson offers options for risk determination that include assessing the organization's ability to implement mitigations or estimating how assumptions relate to what is actually known about a system. She also refers to the Navy's SUBSAFE program and its use of objective quality evidence, "defined as any statement or fact, either quantitative or qualitative, pertaining to the quality of a product…based on observations, measurements, or tests *that can be verified* [emphasis Leveson]" [5, p. 452]. Determining a method for proposing overall risk is an effort left for future work.

*STPA-Based Planning and Inherent System Safety*

A test enterprise traditionally publishes a standard completion report for every project covering the results of the test-objective evaluations. This information contributes to a product's lifecycle documentation as it transitions to the field. As discussed earlier, the test objectives for a particular project can be safety-related in certain circumstances. However, in cases where objectives are mostly technical and major issues are found during testing that are deemed to potentially affect safety in the field, the method to quickly deliver that information to stakeholders is through problem reports. Problem reports document flaws

[110] I was in no position of knowledge to presume the system would behave any differently than was originally documented.

with system performance or safety not anticipated by the stage(s) before the one submitting the report. The updated organizational control structure in Figure 4-7 shows that operations in both testing and the field are capable of providing these types of reports about the system.

During the writing of the notional STPA-based plan for the wingman project, I noted a few issues I found doing my hazard analysis that would impact the safety of the fielded system. This is not to say that the traditional test-safety process would not catch some of these issues. However, they were not component failures or egregious design inadequacies. This means that it would be very unlikely for the traditional problem-reporting method (deficiency reports) to be the appropriate way to document them. Whenever testers write a DR, it must usually be tied to a severe component design problem or component reliability problem. The more insidious issues cannot de documented consistently by that method. More than likely, the issues I found would be mentioned somewhere in the project-completion report if traditional practices were followed.

The concern with this form of documentation is that stakeholders are much more likely to read problem or deficiency reports than pick apart every line of a project-completion report. If something in the completion report does not directly address test objectives,[111] it is not likely to be examined. If engineers in the test stage feel compelled to highlight issues that might affect safety in the field, there should be a consistent method to do so regardless of how severe a design issue it is or whether or not it is a component failure problem.[112] The traditional test-safety planning method does not provide this ability. However, the STPA-based format does. By design, if STAMP is the accident model used for system safety, then test safety is just a special case of system safety. The way findings are published in a test-safety report can contribute to the shared safety-control model of a product across its lifecycle stages. Issues found during the test stage that would affect field use are easily reportable with this paradigm.

One of the issues I noted in the wingman project that might affect field use was in the data-entry interface for the lead-aircraft pilot's requests to the wingman (CA-11b).[113] The method of issuing push-button commands was encumbered by the poor layout of the display and switches that were in an engineering-development format. Several UCAs existed in this control relationship, and causal scenarios involving an incorrect push-button request due to the poor display format were possible. Control 11b is not a test-specific control relationship. In the field, pilots would be expected to send requests to the UV wingman in the same manner. Unless the designers already documented that the display panel will be updated before fielding, testers should report the problem so that it remains in the lifecycle system-safety documentation.[114]

---

[111] One of the test objectives in the wingman project was to address the operational utility of the system, so it is possible that the issues I found would be discussed there. However, soft test objectives that are met only with pilot comments will not consistently report system flaws.

[112] This addresses the fourth research gap discussed in this chapter.

[113] All safety-control structure call-outs can be referenced with Figure 4-12 and Table 4-3.

[114] On the other hand, the similar issue discussed earlier regarding the fourth zero place on the surrogate evaluator pilot's data-entry display would not be reported as a field problem. This is because the evaluator pilot does not exist in the field.

Other issues, alluded to earlier, involved the lack of salient feedback to the lead pilot, through cockpit displays, that requests were accepted by the UV (CC-4, FB-11). Due to affordance feedback, the lead pilot might act on the assumption that the wingman is doing what he intended it to do after push-sending the request. Similarly, salient feedback about datalink dropouts did not exist. The lead pilot would have to know to visually observe for the wingman's motion (CC-3a), if such a viewing angle were available, to confirm the request was received.

The STPA-based test-safety planning format identifies the UCAs and causal scenarios corresponding to the above noted flaws, and the documented hazard analyses can be maintained as part of the larger system-safety database for the product. Testers can easily determine which UCAs are test-specific and which ones would also apply to the field. This level of consistency and rigor—and the ability to document system properties that affect its inherent safety— is not available with the traditional planning format.

*Note on Mission Briefings*

Test-safety planning as demonstrated here focuses on the test-operating phase. In performing the hazard analysis, planners must consider the other phases of test conduct such as maintenance and mission planning. Although aspects of these phases might not be represented in the safety-control structure, they are incorporated in the safety mitigations via inputs like developing influences and settings/objectives. The information conveyed during a mission briefing is usually the last setting/configuration that affects the entire human test-conduct team—the aircrew, engineers, and control-room directors who will actively participate in an operation. [115] Not every industry uses mission briefings before every operating phase. For those that do, it is a good opportunity to align mental models and emphasize aspects of safety planning.

In every mission briefing, safety should be covered in a dedicated portion of the session. The traditional safety-planning format at Edwards provides language in the form of system restrictions, GMPs, and THAs for briefings. My safety study stopped after the notional STPA-based safety plan was produced; I did not go so far as to attempt a mock mission briefing. However, similar to the traditional document's outputs, STPA hazards, MPs, CTAs, RAs, and temporary and permanent system restrictions should be discussed in mission briefings. If local management has put additional restrictions on the airspace, ranges, or other aspects of the test framework, those should be briefed as well as they are really restrictions on the SDT.

It is not easy to brief safety items to a captive audience. During fact-finding interviews I learned that many people become distracted or uninterested during the safety portions of briefings. This is especially the case when multiple mission briefings for the same test project have already discussed the same information repeatedly, or when multiple THAs with repeating minimizing procedures are included in each briefing. Nothing is recommended to combat the repetition of mission briefings, but the repetition of

[115] Analogously, maintenance might be the last setting/configuration affecting the non-human aspects of the system.

information *in* the briefings is avoidable if the STPA-based products are briefed. This format comes at the compromise of MPs being organized by temporality (developing influences, settings/configurations, and operating procedures) instead of tied to unique chains of events as in THAs.

### 4.3.5 Comparison of Methods by a Human Research Study

In addition to discussing and comparing the two safety-planning methods objectively, I administered a survey at Edwards to collect insights from test professionals and compare the planning outputs for the wingman project subjectively. The study protocol was per the MIT Committee on the Use of Humans as Experiential Participants, number 1505697227.[116] The purpose of this study was to determine which of the two safety-planning methods produced a more preferred document in terms of being more intelligible, informative, and implementable. *Intelligibility* referred to the accessibility of information in the document, the ease of comprehending that information, and the intuitiveness of how the information was presented in the structure of the document. *Informativeness* referred to the document's ability to convey information about hazards, the causal scenarios that might contribute to the hazards, and safety mitigations. *Implementability* referred to the ease and willingness of planners to construct (or modify for use) new diagrams,[117] ease of identifying hazards, causal scenarios, and mitigations, and perceived ability to brief, implement, and track risk mitigation strategies.

Comparing the safety plans using these three assessment types allowed for a more granular understanding for why one method was more preferred by participants than another, and whether there are potential trade-offs to be made in selecting one safety-planning methodology over another. In addition to capturing whether STPA would provide additional value above and beyond traditional safety planning in terms of its information, intelligibility, and ease to implement, the survey was also designed to solicit information from test professionals regarding how the STPA-based method might be improved and whether there might be difficulties or barriers in transitioning STPA into the USAF testing community.

#### *4.3.5.1 Methods of the Study*

*Participants*

Twelve participants initially volunteered for the study, and eight completed it. Participants were all 412TW flight-test professionals, responsible for planning test projects and executing test missions from aircraft and control rooms. Some members from the TPS-student team that wrote the traditional safety plan for the wingman project were available to participate. The remaining volunteers were recruited through a call for participation made to several offices under the 412TW. Every participant in the sample is a representative of the key population; however, due to the sampling by convenience there

[116] This study was separate from the fact-finding interviews I had performed during earlier visits to Edwards.
[117] For example, the safety-control structure or system-modes diagram.

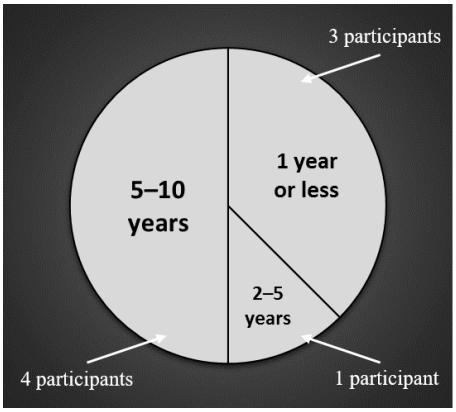



Figure 4-15. Distribution of Test Experience for Participants

was no way within the small sample to proportionally match attributes of the population such as longevity and types of test experiences.

The demographic by test experience of the eight participants that completed the study is shown in Figure 4-15. Three participants had one year or less of test experience.[118] Nobody had more than one year and up to two years. One participant had more than two years and up to five years. The remaining four participants had more than five years and less than ten years. Nobody had ten years or more.

*Procedures*

I administered all procedures of the study. These included an introductory information session followed by the administering of a survey. No personally identifiable information was collected from any of the volunteers during the sessions or in the subsequent surveys. They were not compensated or given any incentives to perform. All participants were told the research had the potential to benefit the AF test enterprise. Participants had the right to withdraw at any point. All provided written consent.

Twelve participants attended the information sessions. Not all were able to attend the same session, due to schedule constraints, so multiple sessions were made available so that each participant could attend one. Each session began with a briefing covering basic concepts of STAMP, STPA, and the format and outputs of a notional STPA-based safety plan.[119] This briefing lasted approximately 60 minutes. Following the briefing, participants received copies of a traditional safety plan and an STPA-based plan for the same project (autonomous wingman). The participants knew I wrote the STPA-based safety plan and were asked not to take that into consideration as they responded later to the survey that would be administered after the information session. Participants viewed each document and were given 30 minutes to clarify anything with me. Each participant was also allowed to contact me in private at any time after the session, although no one did.

[118] Those three participants were TPS students on the team that wrote the traditional safety plan for the wingman project.
[119] All participants were familiar with the traditional format.

Table 4-7. Survey Questions

| Question | Response |
|---|---|
| **Which of the Safety Plans did you find MOST Intelligible?** | **TWO choices** |
| Easy to quickly reference desired information | FOUR choices |
| Easy to read and comprehend | FOUR choices |
| Easy to find the "bottom line" | FOUR choices |
| Consistency of formatting across multiple similar entries (e.g., hazardous behaviors) | FOUR choices |
| Easier to mentally visualize the system | FOUR choices |
| Easy to understand what portions of the system are upgraded / being evaluated | FOUR choices |
| Easy to understand which equipment and personnel are part of only the testing (but not the intended field use) | FOUR choices |
| **Which of the Safety Plans did you find MOST Informative?** | **TWO choices** |
| Informative presentation of hazards (and unsafe actions, if applicable) | FOUR choices |
| Informative presentation of causes / causal scenarios | FOUR choices |
| Informative presentation of minimizing procedures / considerations | FOUR choices |
| Informative presentation of corrective actions | FOUR choices |
| Traceability of causes / causal scenarios to hazards / behaviors | FOUR choices |
| Traceability of minimizing procedures / considerations to causes / causal scenarios | FOUR choices |
| **Which of the Safety Plans would you consider the MOST Implementable?** | **TWO choices** |
| Ease of performing the hazard analysis | FOUR choices |
| Ease of constructing the safety plan document | FOUR choices |
| Ability for the format and information in the document to be used as a template for future documents | FOUR choices |
| Easy to teach the method to someone | FOUR choices |
| Perceived ability of analysis outputs to inform risk mitigation activities during test planning | FOUR choices |
| Perceived ability of analysis outputs to aid pre-mission briefs | FOUR choices |
| Perceived ability to implement changes to the safety planning as lessons are learned during test activites | FOUR choices |
| What do you like the best about each method? | Short answer |
| What do you like the least about each? | Short answer |
| How much time would you recommend to someone for learning the basics of each? | Short answer |
| Which method would you prefer to use for your next test project, and why? | Short answer |
| Do you have any suggestions for the formatting and information ordering in the STPA planning document? | Short answer |
| Additional Comments | Short answer |

After participating in the sessions, participants were given the survey, which was administered online; participants were allowed to respond at their leisure and encouraged to do so. Of the twelve initial participants, eight completed their survey. Besides collecting basic participant demographics, the survey contained two types of questions: multiple choice and short answer. The multiple-choice questions allowed for quantitative analysis of preferences. The short-answer questions were used to provide a means for participants to clarify the reasoning for their preferences and elaborate on criticisms they had of either safety-planning method. Participants were also encouraged to comment on how the STPA-based document could be better formatted and organized. The questions are listed in Table 4-7. They were written to be as unambiguous as possible, cover information available in the provided documents, and—in the case of multiple-choice questions—focus the participants' responses on each assessment type being interrogated [307]. Appendix C contains the raw survey responses, tabulated values, and statistical calculations to support the remainder of this discussion.

*Survey Design Considerations*

As the sample size of participants was anticipated to be small,[120] it was not feasible to design multiple-choice questions aimed at creating a scaled score for each assessment type (intelligibility, informativeness, and implementability) within each safety method.[121] I instead designed the multiple-choice questions to ask for preferences between the two safety-planning methods over multiple fields addressing aspects of the three assessment types. By building distributions of categorical responses, non-parametric methods could be used to look for any deviations away from an equal preference between the safety-planning formats. Although there is no statistical power available to quantify preferences, trends can still be acknowledged and substantiated with short-answer questions.

The blue fields represent a single forced-choice question asked for each assessment type; the participants had to choose which of the two safety plans was most characteristic of that assessment type[122] (i.e., only two possible choices). Immediately after each forced-choice question, several detailed questions (green fields in the table) followed to qualify their responses. For the detailed questions, participants could select between "traditional plan", "STPA plan", "both equivalent", or "neither effective" (i.e., four possible choices). Furthermore, although the three assessment types were always presented in the order shown in the table, the detailed questions that followed each forced-choice question were presented to every participant in a different random order to minimize order effects.

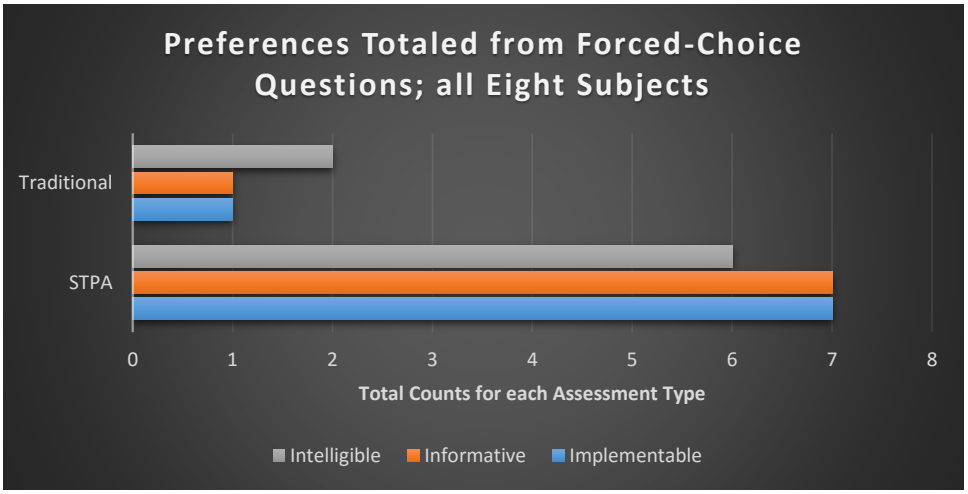
#### 4.3.5.2 Multiple Choice Results

Figure 4-16 plots the observed preferences for the three forced-choice questions as histograms. Raw counts are displayed. The top chart shows the results for the entire participant pool, while the bottom chart divides the pool into the three participants with one year or less of experience and the five participants with more experience. Immediately noticeable is a strong preference for the STPA-based method when forced choices are made for each assessment type. All the less-experienced participants preferred STPA across all assessment types, and this was surprising as they are the ones who wrote the traditional test plan. Some of the more experienced participants preferred the traditional method in some areas; the amount of relative preference toward the traditional method was almost consistent across all three assessment types with intelligibility containing one more count than the other two.

Figure 4-17 plots the observed preferences for the twenty detailed questions. Instead of raw counts for each of the detailed questions, the histograms show the proportion of summed preferences for all detailed questions within each assessment type. Similarly to the previous figure, the top chart shows the results for the entire participant pool, while the bottom chart divides the pool into the three participants with one year or less of experience

[120] I had to accept there might be as few as three participants.
[121] For example, multiple Likert ratings from each participant within each assessment type could be used to develop distributions. Their means and standard deviations that could be used to make parametric comparisons between distribution pairs.
[122] Intelligibility, informativeness, and implementability were defined for the participants in the same manner as the opening paragraph of this discussion.

(a)




(b)

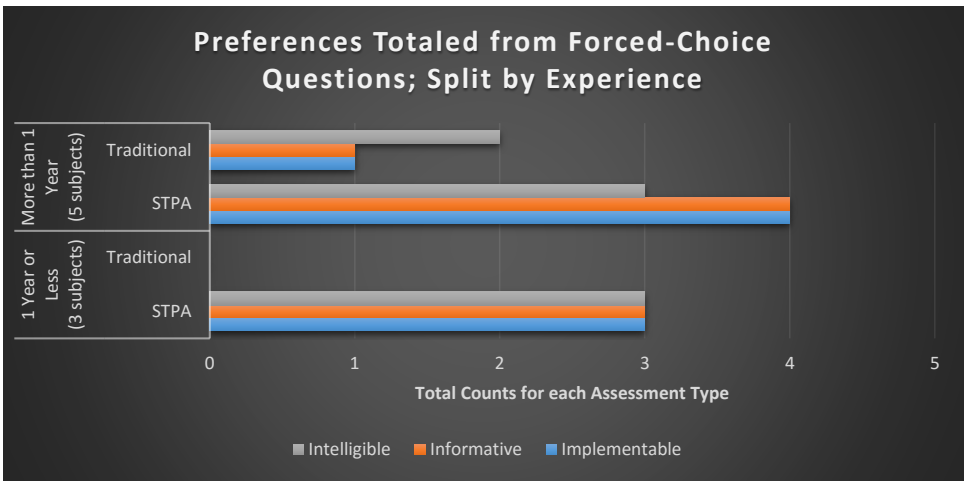



Figure 4-16. Histograms: Forced Choice Questions

and the five participants with more experience. A strong preference for the STPA-based method is readily apparent. However, the distribution of preferences is more varied when examining the more detailed, qualifying responses.

Participants in the less experienced pool were almost as likely to consider both plans equivalently intelligible than STPA alone. However, nobody from the less experienced pool preferred the traditional method alone in the detailed (or the forced-choice) questions. Participants in the more experienced pool responded with less preference for STPA regarding its implementability than less experienced participants. For them the traditional method was almost as likely as STPA to be considered solely more implementable, with some preferring both. However, they showed a heavy preference toward the STPA-based method for intelligibility and informativeness.

In order to quantitatively assess participants' preferences, chi-square tests (non-parametric) were used on the multiple-choice responses. A goodness-of-fit test was administered for each forced-choice and detailed question (see Appendix C). The purpose of the goodness-of-fit test is to determine if a distribution is inconsistent with an expected

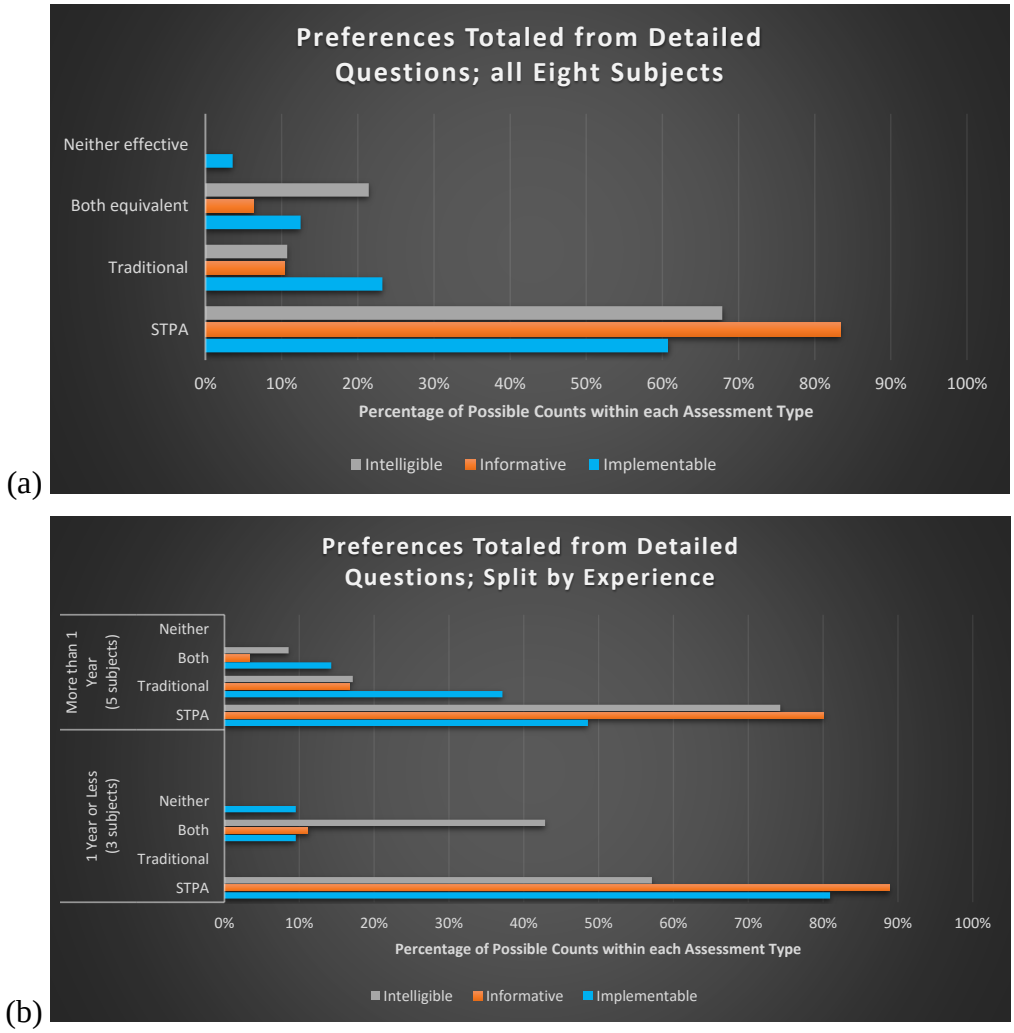



Figure 4-17. Histograms: Detailed Questions

model. For the case of the forced-choice-question responses, the expectation was a uniform distribution of preference. Therefore, the null and alternate hypotheses were:

$H_0$: Responses have a proportionally uniform distribution of [0.5, 0.5] across the two choices

$H_A$: Responses *do not* have a proportionally uniform distribution of [0.5, 0.5] across the two choices

For the forced-choice-question responses, the null hypothesis was rejected for both informativeness and implementability.[123] The test failed to reject the null for intelligibility.

The goodness-of-fit hypotheses for the detailed questions were slightly different than for the forced-choice questions. The expectation of equal preference between the two

[123] Intelligibility: $\chi^2$ (1, n = 8) = 2.00, p = 0.157; Informativeness: $\chi^2$ (1, n = 8) = 4.50, p < 0.05; Implementability: $\chi^2$ (1, n = 8) = 4.50, p < 0.05

safety plans was interpreted across the four detailed choices as follows: of eight participants, the expected distribution would be that three would choose STPA, three would choose traditional, one would choose "both equivalent", and one would choose "neither effective". This was because I briefed the participants to make the best attempt to choose one safety method or the other, and if after careful consideration they chose both or neither, to please add statements in the short-answer responses clarifying what they liked and disliked about each method. Therefore, the null and alternate hypotheses for the detailed-question responses were:

$H_0$: Responses have a proportional distribution of [0.375, 0.375, 0.125, 0.125] across the four choices

$H_A$: Responses *do not* have a proportional distribution of [0.375, 0.375, 0.125, 0.125] across the four choices

For the detailed-question responses, findings varied. For intelligibility, the responses to three of seven questions rejected the null. For informativeness, the responses to four of six questions rejected the null. For implementability, the responses to two of seven questions rejected the null. Therefore, out of all of the responses to detailed questions, nine of twenty rejected the null.[124]

Despite this, failure to reject for any given question was not an indication that the traditional method was more preferred than STPA. The respondents' preferences trended toward STPA for nineteen of the twenty questions, and there were equal counts for traditional and STPA for one question. The nine questions that rejected the null highlighted those detailed aspects in which there was a systematic preference evident. Thus, when clear preferences *were* expressed, they were all categorically directed toward STPA.

#### *4.3.5.3 Short Answer Results*

The raw responses to the short-answer fields are listed in Appendix C. This section paraphrases the information from those comments and relates the short-answer information to the multiple-choice data analysis.

Briefly revisiting the multiple-choice, the nine detailed questions in which participants' responses showed a systematic preference for STPA were (paraphrased):

- Consistency of formatting
- Easy to understand what portions of the system are being evaluated
- Easy to understand what portions of the system are part of the test framework
- Hazards clear
- Hazardous behavior (UCAs) clear
- Traceability between hazardous behavior and hazards

[124] For an organized list of the test statistics and p-values of the responses to detailed questions, refer to Appendix C.

- Traceability between mitigations and hazards
- Ability to use the method to identify mitigations
- Ability to implement changes to safety plan as lessons are learned during testing

Examining the other eleven detailed multiple-choice questions in which responses did not indicate a clear multiple-choice preference, the following seven are paraphrased because they cover aspects that were later addressed by respondents with short answers:

- Ease of referencing information
- Ease of reading and comprehension
- Ease of visualizing the system
- Ease of performing the hazard analysis
- Ease of constructing the safety-plan document
- Ability to use old documents as templates for future ones
- Ability of analysis outputs to aid mission briefing

Moving on to the short answers, it is first useful to summarize what respondents liked the best and the least about each method:

STPA ADVANTAGES

- It identifies contributions to hazards inherent in the entire system (not just items under test); better for determining true risk
- Description of the system and boundary are more accurate and explicit, and the distinction between accidents and hazards is clearer
- The structure is more straightforward and easy to follow, and traceability of hazardous behaviors and mitigations is built-in

STPA DRAWBACKS

- It requires an intricate control analysis and more time to perform appropriately
- It can be difficult to navigate for larger projects with a wealth of information, especially with the traceability expressed as parenthetical references
- It requires more management involvement in terms of system definition, standardization of terms and formats, maintenance of repositories, and teaching of the new method.

TRADITIONAL ADVANTAGES

- It is more familiar and hence more comfortable
- It is fast and convenient, especially from reusing of old planning documents to aid in writing new ones

- Test hazards, as defined traditionally, are test-specific and easy to brief and keep in mind during a mission
- Easier for decision-makers to visualize test-specific hazards and qualify risk

TRADITIONAL DRAWBACKS

- It encourages laziness in the analysis without a full understanding of the system, due to the ease of copying old safety plans as well as duplicating test-hazard sheets and mitigating procedures
- It relies on experienced reviewers to catch any holes that were missed by planners
- During mission briefings, repeated reviews of multiple test-hazard sheets with overlapping information tends to cause practitioners to tune the information out
- It is unclear what belongs in the technical plan and then the safety plan, often resulting in repeated information in both

One commenter warned that either method could over-constrain the test team by placing undue mitigations on the test activities before the system during test is better understood. This is a valid concern that emphasizes the need for both a common model of the system starting in the design stage, as well as the documentation of design and use assumptions within that model. If such a paradigm were in place, the STPA-based method would very easily be able to help test planners determine the best areas to apply intentional constraints during the test activities, and any mitigation deemed to be too constraining could be examined with respect to the safety-control structure to search for alternative solutions.

Revisiting the data plots, based on forced-choice questions the experienced participants trended slightly more away from STPA in intelligibility than the other two assessment types. Based on detailed questions, the less-experienced testers trended "both equivalent" almost as much as STPA in intelligibility. Relating that observation to the comments, participants may have found difficulty in navigating a large magnitude of hazard-analysis data. The STPA-based plan I gave the participants contained an example of the worded list of UCAs and causal scenarios that might accompany each type of test maneuver discussed in Section III of the document. It was also not very easy to use parenthetical traceability between safety mitigations at the end of Section III and UCAs/causal scenarios.[125] A solution to these issues would be to contain the STPA Step 1 and 2 details in a separate but accessible spreadsheet or database that can be linked to by the safety plan. This concept would require reconfiguring the test wing to a new form of safety planning that does not expect the product to be a single text-written document. That highlights the other issue seen in the visual trends: implementability.

Experienced users were almost as likely to trend toward the traditional method as they were STPA on implementability based on the detailed questions. The consensus among those that commented further was that the STPA-based method requires significant management involvement to standardize the definitions and boundaries of the systems

[125] I also attempted to use parenthetical comments when writing the SDT description, to provide cross references to the safety control structure. One participant specifically commented that it was detracting.

during test, maintain coordinated repositories for accidents, hazards, and traceable databases, and implement and teach the new method. When asked to comment on how long it would take to teach STPA to fellow testers, one participant likened it to teaching someone to use a graphing calculator over a four-function solar calculator; it might take longer to do, but it is worth the investment. Other opinions estimated both methods as taking hours to months to years to not only be taught well but learned tacitly through experience. One participant recommended that safety-office personnel have a systems background to begin with if the STPA-based method were to be implemented.

Furthermore, an STPA analysis for a specific test project would, regardless of how well the practitioners know the method, take longer to accomplish in an already schedule-constrained enterprise.[126] This argument, however, underscores one of the primary benefits of the STPA-based method: it forces more thought to be put into a true systems analysis. This would require some consideration within the enterprise as to what matters between the competing priorities of getting a safety plan approved expediently and comprehending the hazardous behaviors possible in newer systems with increasing complexity.

With the small sample size of participants—all selected by convenience—it was not possible to control for participant predisposition or apprehension, except to encourage the participants to be open and honest without worry any retribution or offending of my product. Responses were assumed to be genuine due to significant amount of personal time each participant volunteered to the study. Demand characteristics were possible in that the participants knew the purpose of the study was to support test modernization, and there was no efficient method available to blind the treatments or divide the time spent in the information sessions more evenly between teaching the traditional method and the STPA-based method. Most of the time was spent discussing the nuances of the STPA-based method. Even less practical was a double blind design, as I was the only one available to administer the study.

Despite these limitations, none of the three less-experienced tester who participated in the survey chose the traditional method for any of the questions, general or detailed. They were all TPS students who wrote the traditional safety plan, and the expectation of some anchoring or availability bias was not substantiated by the evidence. Although it is possible all three of them were displaying good-participant characteristics, I made it clear during the sessions that there was no right answer and that dissent was encouraged for all choices and comments.

The five experienced testers may have had a bias contributing to a noticeable amount of their detailed implementability preferences favoring the traditional method; this would not be surprising given their availability of practitioner knowledge and predisposition and comfort with the traditional techniques. The short-answer comments discussed provided the needed insight into what would be required to consider instituting a modern safety-planning method like STPA.

---

[126] The current traditional method is not really a formal hazard analysis (i.e., fault trees), so any formal method like STPA will naturally take a little more time to accomplish than what exists today.

When asked to provide an optional short-answer comment stating which method the participant preferred for his or her next test project, six participants answered with five choosing STPA. Those supporting STPA (two TPS students and three more-experienced testers) said the traditional method is not taught very well to begin with, the larger up-front investment is worth it to get safety-planning right, and the clear distinction between accidents and hazards sets up the analysis to be easier. The participant who chose the traditional method (more-experienced tester) did so because he or she knows the process well and could perform it in a timelier manner; however, that participant also admitted he or she would want access to the type of information contained in the STPA analysis.

#### *4.3.5.4 Conclusions of the Study*

In summary, there was a systematically detectable preference for STPA in the responses for two of the three forced-choice questions and nine of the twenty detailed questions. The responses for the other questions tended to favor STPA, but failed to create a detectable pattern away from the assumption of equal preference. With only eight survey participants, this is still useful data. However, in order to *measure* the difference in preferences, a study with a larger sample size and parametric methods would be required. Rating scales could be used to score both methods for each question to facilitate a within-participants design for analysis of variance.

Transitioning the safety-planning process in AFTC requires a thorough sketching out of the issues emphasized in this study over more specific examples within the test community. Based on the comments in the survey, a future study should revisit the manner in which hazard-analysis data are organized and accessed in the STPA-based safety plan—avoiding document wordiness and parenthetical references.[127] More advanced information-management techniques might be warranted. STPA Step 1 and 2 details could be maintained in a separate but accessible spreadsheet or database. Hyperlinks or footnotes could be provided as each test maneuver is described in the safety plan to be directed to the applicable UCAs. Because MPs are already numbered in the safety-plan document, the databases could list them against UCAs (which themselves already trace to hazards). This enables the option for the reader to reference the hazard analysis without cluttering the written plan and making it unintelligible.

The future study should also be conducted on a test project that can be analyzed with the cooperation of experienced test planners knowledgeable in the technical background of the candidate system. Ideally, design-stage practitioners would also be available to clarify details about the system. The resulting safety plan should be reviewed by the same type of expert panel that reviews a traditional safety plan. This would help establish guidelines for how an STPA-based plan goes through the review and approval process and would help clear up any common misunderstandings, inefficiencies, and inconsistencies that come about from using new format.

---

[127] One comment also recommended mitigations be kept to a single page if possible, at least when presented for management approval or for test mission briefings.

Twice in the comments, the prospect of defining a better way to divide the information requirements between a technical plan and a safety plan was addressed. Taking this idea an evolution further, test practitioners should consider that writing a distinct technical plan and safety plan might result in wasted efficiency. Writing a safety plan after a technical plan[128] is also reductionist because in practice it applies safety mitigations as afterthoughts onto chains of component events, after the test configurations and procedures have been planned on an existing design. With STPA, hazards, the safety-control structure, modes, and worst-case emergent behaviors can be acknowledged for the SDT from the beginning of the planning. An integrated test-and-safety plan that defines and documents these aspects as the techniques and procedures of the technical strategy are discussed and developed could benefit the tracking and mitigation of UCAs. This concept would be a significant paradigm shift and require further discussion in AFTC.

Implementing STPA requires focused effort, both in the transitioning of the enterprise to a new method and in day-to-day safety planning and reviewing. The decision to consider it for further review is one for AFTC stakeholders. Furthermore, the bigger picture of standardization and continuity between the design, test, and fielding stages needs to be addressed in any organization desiring a systems view of testing. The knowledge that designers have about the system cannot be easily replicated by testers, and the testers also need to understand how the system will be used so that they can operate it in manners that evaluate the functions it will be asked to perform in the field.

## 4.4 Value Added to Test

Developmental test is typically where system design meets operation for the first time. The most targeted and highly scrutinized safety-review of a product—led by highly skilled and experienced discipline engineers and operators—happens in the test stage.[129] This stage of the product lifecycle can take advantage of this extraordinary concentration of keen planners and evaluators to not only plan for test safety but report on inherent system safety in a detailed and efficient manner.

A generalized example of a STAMP organizational control structure with a test stage did not exist before this research. Updating it to foster a systems view of testing has provided several benefits for testers and system managers. The socio-organizational aspects of a test enterprise can now be considered and modeled using new inclusion criteria. The types of communications and reports that the test stage shares with other stages of product development may be managed by stakeholders. The test enterprise may begin to model aspects of its operating process that are common to the testing of many different systems. These test-framework items combine with new and legacy components of a system to form a complete SDT when a new product arrives to the test stage.

Using STPA to perform test-safety planning allows for a proactive, worst-case analysis of the SDT that does not rely merely on hindsight for planning and is more robust

[128] Although it is encouraged for planners to write the technical and safety plans concurrently, based on fact-finding interviews at Edwards this almost never happens in practice.
[129] As mentioned earlier in Section 4.2.2.2, specific policies exist in the AF governing the test-safety practice.

to the inconsistent expert knowledge that might be available during safety reviews. When emerging technologies have no experience or expertise basis on which to plan, the ability to visualize the system, understand the hazards, and perform a systematic analysis that is also system-theoretic is an advantage. The descriptions of the system and environment are explicit, accidents and hazards are clear, and traceability of hazardous behaviors is natural. Safety mitigations do not repeat themselves and are not created solely by copying old safety plans, and there exists the potential to aid technical planning as well by assuming a systems view from the beginning of a project.

The improvements to test-safety planning alone are akin to many documented successes of STPA applied in various industries worldwide. What makes the systems view of test immensely useful is the recognition that test safety is a special case of system safety. Traditionally, the findings of test-safety planning inform only the test activities, unless a defect is found so severe that a problem report can be submitted to immediately fix the design or manufacturing process. Smaller issues become lost in standard test-completion reports, and in the best case testers may be able to warn field users to be aware of the issues or try to write a problem report that does not explain system impacts through any means but a written narrative. However, with the systems approach, the attention to detail that has gone into test-safety planning for the SDT is already informing aspects of the system as intended for fielding. Hazardous behaviors found during test planning can come from any part of the system, not just the test framework.

With STPA, test engineers can *manage risk for local test events* while concurrently analyzing the design and use of the system for inherent system safety. Undesirable behavior and design flaws can be traced in the hierarchical control structure and documented in a model of the system that is shared between lifecycle stages. This is particularly useful for software and human experts who must be able to consider field-use implications while managing the knowledge gaps in the design. STPA-RC introduced refinements that enable engineers to elucidate aspects of the system particular to intelligent controllers. Furthermore, developing influences and settings/objectives can be considered when identifying hazardous scenarios, and they are also explicitly included in the safety-planning format as types of safety mitigations.

The ability to emphasize what hazardous scenarios are specific to test while also contributing to the system-safety control model of a product is a powerful contribution of STPA to developmental test. Furthermore, modeling a test stage using STAMP provides industries that manage system safety a template for developing the safety-planning considerations their testers should adopt in order to evaluate a product while ensuring the risks of test activities are managed. With details modified where necessary, the new organizational control structure example, inclusion criteria, test-safety planning sequence, and planning-document format are applicable to any industry. In its current form as of the writing of this thesis, the entire process may be used on any defense or commercial development of an aircraft or weapons system.

# Chapter 5

# Conclusions

> "You look at where you're going and where you are and it never makes sense, but then you look back at where you've been and a pattern seems to emerge."[1]
>
> —*Robert Pirsig*

The purpose of this research was to:

> *Improve the ability to assess system safety during developmental product testing and standardize the applicability of hazard findings between the design and field use of the product.*

Its two complimentary objectives, discussed in Chapters 3 and 4, respectively, were:

1. Extend STPA to better examine human controllers in the hazard analysis.
2. Provide a common framework for test-safety planning that addresses both the safety of the test process and inherent system safety.

## 5.1 Summary of Work

In order to satisfy the first research objective, this thesis developed an extension called STPA-RC that refines the analysis of human controllers. Former analyses were updated to incorporate existing models and taxonomies and increase granularity of the controller analysis by including new guidance across various intrinsic and extrinsic factors that contribute to unsafe behavior. The analysis was made general to both human and non-human controllers, with portions specific to only humans appropriately designated, and with new types of system abstraction now included in the process model to account for intelligent controllers.

The extrinsic factor of *influence* was created in order to capture controls that begin to develop prior to an operating process. This allows the analyst to focus the hazard analysis on a specific phase; the controls and settings from previous phases or higher entities in the organizational control structure that affect the actions of the system's controllers during the phase can be considered rigorously with the additional guidance provided. Influences

[1] *Zen and the Art of Motorcycle Maintenance* [308, pp. 167–168]

can also be considered when exploring viable options for designing hazard mitigations. Visual aids were developed to assist both the main STPA-RC analysis and the additional guidance for influences.

An Explicit-Influence Map was proposed as a visual planning tool that charts the various standards, rules, procedures, and provisional settings identified across a sociotechnical enterprise that flow to front-line operators. This tool gives practitioners and stakeholders a means to quickly visualize the relationships between items and identify any gaps, redundancies, or conflicts that could contribute to inappropriate controller behavior. The ability to visualize and discuss inadequacies with policy can aid practitioners in the ability to recommend improvements in those documents and identify appropriate instances when they should codify otherwise tacit standards and practices.

In order to satisfy the research objective, STPA was tailored to perform test-safety analysis. This was important because test safety requires a unique set of best practices and planning activities that are required in system-development organizations because of the distinctive configurations, modifications, measurements, and techniques required to evaluate products. Taking inspiration from the Air Force acquisition model, the generalized example of a STAMP organizational control structure was updated from Leveson's original format (itself inspired by Rasmussen) to add a testing stage to a product lifecycle. Although the new example includes a separate organizational hierarchy for each stage, different industries might have varying control structures to match each of their particular bureaucracies. By visualizing a discrete test stage however, the information that flows between designers, testers and field users is highlighted on the new figure, as well as the test-specific safety communications that exist during product evaluation.

The concept of developmental stages was also used to propose that a system-safety control model and analysis for every product should be developed by designers and shared with testers and users. The ability for practitioners at various stages of product development to contribute findings and problem reports to the shared analysis contributes to ongoing system-safety certification. Testers also assume any increased risk of operating immature systems so that the specifications they verify and knowledge gaps they address will contribute to updating the system hazard analysis.

With a shared control model and analysis, testers can create a product-specific control structure that includes the test framework. Any unsafe behavior found during the hazard analysis that is due to elements of the test framework or in the procedures used to exercise the aspects of the system being evaluated are test-safety considerations, making test safety a special case of system safety. Additionally, unsafe behavior found during the hazard analysis that is related to the items under evaluation, regardless of what level of the system the behavior emerges, can be used to inform the shared analysis of the product.

An Air Force product test was studied for use as an example of performing STPA both to address test-safety planning and to improve the inherent safety determinations of the product. The many safety programs and practices in the Air Force were discussed and visualized within an Explicit-Influence Map that also included general acquisitions, test, and operations policies. Then, the traditional method for test-safety planning was presented and used as an inspiration to develop a template for a proposed STPA-based planning

method. The traditional and STPA-based methods were compared in general, followed by a summary of the test-planning documents each method produced for the Air Force product.

Air Force test practitioners wrote the traditional document, while I wrote the STPA-based planning document. Objective results were compared to include the numbers and types of safety mitigations identified and/or recommended by each. I additionally found unsafe behavior that would impact inherent system safety in the field, and I discussed examples. The documents were also compared subjectively by means of a survey administered to eight test practitioners that included some of the personnel who wrote the traditional report. Their responses were analyzed statistically. Along almost half of the measures, a systematically detectable preference for STPA was found. The other measures appeared to favor STPA, but the sample size was too small to detect a systematic departure from equal preference.

### 5.1.1 Contributions

The following is a summary of the contributions of this thesis by research gap:

STPA-RC

A) The *process model*, one of the main concepts of STAMP, does not capture types of system abstraction that human controllers require to contribute adaptivity to system behavior.

   The guidance for the controller was refined to examine the consistency, completeness, and correctness of not just the process model—as STPA originally recommends—but also the controller's model of three types of system abstraction. These are behavior, modes, and motives. Human controllers understand and manage these aspects of a system.

   STPA-RC prompts analysts to consider the feedback that designers anticipated as well as feedback that human controllers adapt over time to use for updating their process models. Affordance feedback was also introduced as a consideration to address the human controller's belief that an action has been commanded properly without receiving feedback from the controlled process.

B) Fundamental human considerations are not explicitly considered in the controller analysis (e.g., workspace factors, variability of personal traits).

   The human considerations of workspace and variability were added to STPA by acknowledging extrinsic factors that were contained in prior human-controller models in the STAMP literature. Those factors were refined in the analysis by introducing new guidance.

C) There is no current method to model the impact of social and organizational influences on the controllers within the operating process.

   The additional extrinsic factor of influence was added to the analysis. By defining and acknowledging the work phases in a system, analysts can focus on the behaviors and control actions within a phase. They can take into account

higher controls and settings that began to develop prior to the phase in time through new guidance that captures culture, standards, rules, and settings/configurations.

The Explicit-Influence Map is a useful tool for both research purposes and for practitioners to understand and improve their organizations' policies. The map I created for the Air Force 412th Test Wing is already being considered by local practitioners for use at a basic capacity for referencing policies.

TEST SAFETY

D) There is inconsistent expert knowledge at any given test-safety review board.

See next.

E) There may be minimal expertise in new technologies (e.g., software, autonomy).

Hindsight is not always available when planning for new systems, and when it is it, past events should not be the sole source of mitigation recommendations. STPA provides a consistent and rigorous method for approaching a hazard analysis with a systems-based accident model, regardless of the backgrounds of those performing it. While expertise in a system is always a benefit when performing any analysis, testers may take advantage of the consistency of the STPA-based approach whether or not deep system knowledge exists. Instead of just looking back at prior experiences, testers should look back at what the designers of a specific system intended through their documented models, knowledge gaps, and traceable specifications.

F) The test-safety planning process does not use common visual aids in its documentation.

STPA requires a model of the functional safety control hierarchy to be created for the system. This in itself serves as a useful visual planning tool that can be shared by engineers, operators, and stakeholders for the common purpose of planning for test safety and inherent system safety. Each person's local understanding of the system can be put into the perspective of the shared testing activity, fostering collaboration on the analysis.

Additional visual aids were developed in this research. STPA-RC, as mentioned earlier, provides diagrams to assist with its new analysis guidance. The newly created Explicit-Influence Map is a universal aid that can be used for many areas of organizational improvement including safety. A modes diagram was also proposed to assist analysts in understanding and discussing system and component modes.

G) Test engineers do not have a consistent method of tracing undesirable behavior or potential design flaws to effects on the system within the context of field use; this especially affects human-engineering experts, who cannot ignore the relationship between operating philosophy and system design.

See next.

H) Problem reports tend to be reductionist (e.g., manufacturing error, component defects) and do not consistently explain system impacts through anything but written narratives.

The use of STPA allows safety findings to be put into the context of the system design and how it is operated. Undesirable behavior and/or design flaws found during the hazard analysis are already traced to a functional safety-control structure. Engineers, informed by test planning activities and test-conduct findings, may generate unsafe causal scenarios that are based on any aspect of the system. These scenarios may involve more than just severe component design failures or reliability problems, which are the only aspects that traditional problem (deficiency) reports can handle. The ability to produce scenarios in the context of functional control is particularly useful for human-engineering experts because they cannot treat operators as simple components in systems.

Safety findings at any stage of product development can be documented appropriately by system managers if a common model of the system is shared between stages. Although the model of the system during test contains unique aspects due to the test framework, testers can still contribute findings to the system as it is intended to be fielded. Using the test-project example in this thesis, I found two unsafe behaviors that would impact inherent system safety in the field; some behaviors might only be possible to identify when testers operate the system for the first time. With STPA, test engineers can develop and document preemptive risk mitigation strategies (or recommend changes to the designers) before the system is fielded. These findings and recommendations can be reported within the context of the system model and may take the form of redesigns, engineering devices, warnings, or procedure changes. The order of precedence for applying mitigations should not be based on the form of the mitigation but on system hazards; recommendations should try to eliminate, reduce the likelihood of, and then reduce exposure to hazards.

I) STPA control structures do not acknowledge an explicit product testing stage to capture the particular sociotechnical dynamics found in a test enterprise.

The STAMP organizational control structure example was updated to propose a test stage and inclusion criteria for practitioners wishing to model their test enterprise. The socio-organizational aspects of a test enterprise are part of the system during test and influence the behavior of controllers during the operating phase. Although the proposed diagram is an example only, with details changed the entire process handles any defense or commercial development of an aircraft or weapon system.

Testing is a distinct form of system operation in that the system during test contains a test framework, and often unique techniques and procedures are used to exercise it. Test-safety documentation is likewise unique, and as mentioned above a planning method and document format were proposed that incorporate STPA, use visual aids, and provide engineers a systemic framework

for performing a hazard analysis. However, because test safety is modeled based on a system-theoretic approach, test safety is itself a special case of system safety, so even if the local planning methods involve unique planning formats and tasks, findings can inform design and field-use mitigations.

### 5.1.2 Limitations

All the work I performed to produce diagrams and conduct analyses did not take advantage of any automated assistance. This added a substantial margin of preparation time for methods that would be frequently and commonly used in actual implementation.

The Explicit-Influence Map currently can only be created and edited manually, and users cannot share live updates or comments with other users, except to send out a new updated version of the complete map. The freeware used to create the map is extremely versatile for visualization and basic document linking, but it lacks the ability to connect to a back end or be networked across an organization.

The safety-control structure diagrams demonstrated in this thesis were also created manually using standard office software. Labels and comments had to be numbered and updated individually, and any time the model had to be updated for any reason, the process took a considerable amount of time.

I performed the hazard analysis of the example test project using manually created tables, which did not automatically link to the safety-control structure. With a dedicated software tool, analysts would be able to select portions of the safety-control structure and be directed to an appropriate database for recording hazard causal scenarios. Safety mitigations could be traced automatically to causal scenarios (and thus hazards and mishaps). The proposed safety-planning document lacks a method to link to such a database, and manual parenthetical references within the document were demonstrated but disliked by local practitioners because they made the document unwieldly.

Under ideal conditions, including having access to pre-existing template and automated planning tools, the STPA-based method would take roughly four times as long as the traditional method to produce a safety plan. This is largely due to the fact that the traditional method chosen for comparison (Air Force test-safety planning process) more closely resembles a Preliminary Hazard Analysis than a formal quantitative hazard analysis. STPA uses mostly the same process regardless of whether it is performed by designers, testers, or field users on a preliminary or detailed design or concept. The level of rigor provided by STPA for the example in this thesis was greater than that of the existing traditional approach. Three-hundred percent more time invested yielded about 330 percent more safety mitigations. Those mitigations were non-repetitive and organized to be more expediently briefed before test missions; this came at the expense of the additional time invested to produce them and the lack of the method identifying individual test-specific hazards as is currently familiar to planners. However, the mitigations and the causal scenarios they address are traceable to a model of the system that can be shared with designers and field users, giving the test-stage safety-planning activities more relevance to the acquisition process.

The human-research study was non-parametric and lacked statistical power. With the small sample size, it was not feasible to design multiple-choice questions aimed at creating a scaled score for each assessment type in order to parametrically compare distributions between the traditional and STPA-based plans. Only preferential tendencies could be discussed, and short-answer comments were used to further investigate some of the trends observed. Systematic preferences detected in the data were highlighted.

Volunteers for the study were recruited by convenience, and while every participant was a representative of the key population there was no way in the sample to proportionally match attributes of the population such as longevity and types of test experiences. It was not possible to control for participant predisposition or apprehension, except to encourage the participants to be open and honest without worry any retribution. Demand characteristics were possible in that the participants knew the purpose of the study was to support test modernization, and there was no efficient method available to blind the treatments or divide the time spent in the information sessions more evenly between teaching the traditional method and the STPA-based method. Even less practical was a double blind design, as I was the only one available to administer the study. The participants knew I wrote the STPA-based safety plan and were asked not to take that into consideration as they responded to the survey.

## 5.2 Recommendations and Future Work

### 5.2.1 STPA-RC Recommendations

I did not explore any particular decision-making theories within the analysis developed in STPA-RC.[2] This is an area suitable for research, especially to incorporate theories concerning teamwork operations. The relationships between detection, process modeling, and decisions that form the human's observation loop would also benefit from further research to advance the analyses performed on intelligent controllers. Decision aids could be further studied as well to assess the best methods to analyze them using STPA.

Within the mental model, no explicit method for mapping motives and identifying how they affect (or are affected by) local controller behavior was chosen. Additionally, although human-specific extrinsic factors such as workspace and variability were outlined for the analyst's consideration, feedback that take the form of non-verbal or affectual communication were not explored in this research to determine how they might manifest through the intrinsic factors of detection and process modeling.

The Explicit-Influence Map should be further developed into an interactive, networkable database. The ability for practitioners to highlight conflicts and gaps in policy should be prioritized, as well as the ability to identify and communicate tacit knowledge that should match explicit knowledge in an organization. The Air Force Military Flight Operations Quality Assurance program, for example, could benefit from a method that

[2] For example, Rasmussen or Kahneman [8], [145].

assists in managing discrepancies between tacit and explicit influences. This can help stakeholders identify improper influences to the operating phases in their industries.

### 5.2.2 Test Safety Recommendations

If STPA were to be implemented for test planning purposes, automated tools and templates for the method would be necessary. The top-down philosophy of STAMP should always be encouraged when teaching the method to engineers, as feedback I received from practitioners indicated a worry that the new method would be too detailed to be accomplished with schedule constraints. A future study should revisit the manner in which hazard-analysis data are organized and accessed in the STPA-based safety plan, emphasizing a capability to link to external databases. This would enable the option for the reader to reference the hazard analysis without cluttering the written plan and making it unintelligible. It would also make the natural traceability of accidents, hazards, scenarios, and mitigations easily accessible.

The best applications to continue examining STPA specifically for flight test are autonomy-capable and/or unmanned systems whose designers are willing to share in the safety-control modeling and hazard analysis tasks with testers. The planning should be conducted by experienced test planners knowledgeable in the technical background of the candidate system. The resulting safety plan should be reviewed by the same type of expert panel that reviews a traditional safety plan. This would help establish guidelines for the review and approval process and fine tune the process for efficiency. Stakeholders will have to agree on a preferred method for determining overall project risk based on the findings of causal scenarios, safety mitigations, and other forms of objective quality evidence [5].

If the systems approach were found to be helpful for safety planning, stakeholders should also consider performing technical planning with the same mindset. A hierarchical control structure and system-modes discussion might be just as appropriate in the technical planning as in the safety planning. The ability to identify contributions to hazards as the technical test planning is being accomplished would negate the need to repeat similar information in a safety planning document, and if modern database and reference techniques are used a combined planning document could be utilized that would not be a burden to comprehend for decision makers.

There are many intricate factors that come into concert during a test. Operating procedures, test-safety considerations, design choices, maintenance actions, organizational pressures, crew coordination, and the operating environment itself all contribute to system safety. No single independent safety discipline (e.g., flight, test, range) can currently account for all possible contributions to hazards. Likewise, accident and close-call investigations have a tendency to highlight traditional event-chain narratives and root cause determinations without looking at dynamic interactions among components of the system. However, this form of risk management is not specific to the detailed system design, the way it is operated by a specific set of users, or the organization that manages the system. Hazards must be examined with respect to the specific system.

The various safety policies that I found in the Air Force exist through disparate regulatory legacies, and for the most part they provide procedures and best practices for designers and operators. For example, crew resource management sets some standards by which controllers communicate and problem solve; operational risk management tries to mitigate general conditions that have statistically been determined to have contributed to past incidents. There can be no realistic recommendation to consolidate some of these practices, but they can be incorporated into a specific system's hazard analysis; their effects on controller behavior can be visualized with controls in a safety control structure and/or with an STPA-RC influence analysis, aided by an Explicit-Influence Map.

In order to address the system's specific design and use, information about the system must be documented and shared among developmental stages. I recommend that the Aircraft Information Program shared by Air Force use commands, safety offices, and system managers also include a safety-control structure.[3] Airworthiness should be considered more broadly to *be* system safety. A system's airworthiness determination should include the analysis of not only its specific design but its use philosophy and operating procedures. To do so properly requires the hazard analysis to be system-theoretic and not just a set of arbitrary line-item certification standards. Both test-completion reports and problem reports should be able to update the common system-safety analysis data. Deficiency reports, mandated by a separate regulatory source and maintained in a separate database, can still be used as intended for component defects. However, testers and field users should be able to report the same issues in a systemic format for the Aircraft Information Program, as well as issues not severe enough to be considered in deficiency reports.

Using STPA for test-safety planning allows for proactive hazard analyses that can overcome the limitations of hindsight bias and inconsistent expert knowledge. The systems view of test recognizes that test safety is a special case of system safety. The detailed analysis that goes into test-safety planning can inform aspects of the system as intended for fielding. Undesirable aspects of design and procedures can be stored in a safety analysis that is shared between lifecycle stages. This is a powerful contribution of STPA to the profession of developmental test. Industries that manage system safety now have a template for developing the safety-planning considerations their testers should adopt in order to evaluate a product while ensuring the risks of test activities are managed. The organizational control structure example, inclusion criteria, information requirements, and safety-planning formats demonstrated here are applicable to any industry with details modified as necessary.

[3] Some commercial research partners are already using STPA to analyze their flight operations quality assurance data.

# Appendix A

# Unmanned Vehicle Accident Data

This information is gathered from Air Force (AF) public records and supports the unmanned vehicle (UV) accident discussion presented in Chapter 2. Data for manned vehicles (MV) are not shown, but they were similarly obtained for any UV/MV comparisons discussed in this thesis. Mishap classes are defined from "A" to "D" in decreasing severity, as outlined in DODI 6055.07 [291, p. 36] and concurrently AFI 91-204 [35, pp. 20–21]. A and B mishaps are considered major, requiring an AF-level safety review. Class A mishaps result in death, permanent total disability, total destruction of an aircraft, and/or greater than $2,000,000 in damage. Class B mishaps result in permanent partial disability, hospitalization of three or more individuals, and/or greater than $500,000 in damage. The monetary thresholds were increased starting in fiscal year (FY) 2009 (before then, Class A was defined at $1,000,000 and Class B at $200,000).

## Aggregate Mishap Data

Tables A-1 through A-3 are derived from the AF Safety Center (AFSEC) incident data repository[1] and present the total flight hours and incident rates for the three large-scale AF UV systems: the MQ-1 Predator, the MQ-9 Reaper, and the RQ-4 Global Hawk.[2] Each table section depicts a type of incident (Class A including destroyed, destroyed only, Class B, or Class A+B) and contains several columns. "#" represents the incident count during the respective FY. "C-#" is the cumulative count from the beginning of service life to the respective FY. "Rate" is the number of incidents per 100,000 flight hours for the respective FY, while "C-Rate" is the cumulative value. "E-Rate" is the aggregate incident rate from the respective FY until the present. This value is useful for quickly retrieving the incident rate over the most recent desired number of years to the present. For example, to determine the Predator's Class A mishap rate for the last ten years, one would check the E-Rate value in table A-1 corresponding to FY05.

It is also possible to manually calculate cumulative mishap rates for any range of years and/or combination of airframes by summing the incident counts within a desired bracket of years and dividing by the difference in cumulative flight hours within the

[1] http://www.afsec.af.mil/organizations/aviation/aircraftstatistics/
[2] The Predator is a Group 4 UAS, while the Reaper and Global Hawk are Group 5 (see Table 2-5).

bracket. Tables A-4 through A-6 present a summary of mishap rates calculated by this method.

The aggregate data are also presented in plots via Figures A-1 through A-4. The figures show the AF Predator, Reaper, Global Hawk, and all combined, respectively. For every FY, the types of mishaps (Class A destroyed: black, Class A not destroyed: red, and Class B: yellow) additively construct bars. The bar sections are stacked so that the height of the black and red sections together represents all Class A mishaps, and the height of all three colored sections together represents all major mishaps. In the background, flying hours per year are represented by a connected outline that is filled in blue to aid in contrasting between mishaps and hours.[3]

Within each plot, the two vertical axes—flying hours and mishaps—are intentionally mated. One mishap count is matched to every 10,000 flight hours. This is done to make possible a visual approximation of each year's local mishap rates without cluttering the plot with more numbers. The height of the flying-hours curve corresponds to the height a bar would reach if it produced 10 mishaps per 100,000 flight hours. If a bar looks to be about half the height of the hours curve, it represents roughly 5 mishaps per 100,000 hours, and so on. Each bar can thus both be referenced against the mishaps axis to determine the exact count of the mishap type (corresponding to the bar) and visually compared to the hours curve to estimate the mishap rate. The previous tables contain the actual mishap-rate values should they be desired.

The plots for the Predator and Reaper use the same global vertical scale so that a visual comparison may be made between those two similar UVs. The Global Hawk, having amassed far fewer flight hours in comparison, is plotted on a smaller vertical scale. The scale is once again adjusted in Figure A-4 to account for the combined flying hours of all three UVs.

All three UV types had a noticeable decrease in mishap rates between the first half and second half of service life, whether the dividing line is considered in terms of years or hours. This is also visually evident in the plots. This trend is similar to that seen in MV aviation safety [22]. It should be noted that the Global Hawk has barely amassed over 100,000 life hours. Although this fact makes the interpretation of early year mishap rates dubious (and yearly mishap counts are in the ones and twos), the Global Hawk has followed the same general trend as the other UVs.

Use of the Reaper is slowly surpassing that of the Predator. The Reaper is considered an improved version of the Predator, which has suffered from many hardware defects [19], [207]. Although the Reaper has experienced less issues due to mechanical problems, its mishap rates have also followed the same trend, and causal factors are attributed more to human-factors (HF) problems. The following section discusses causal factors.

[3] Although the blue shaded area looks like a density function, it is not. The plot axes are discrete. Visually, however, this format for presenting flying hours does construct a mental image of how the year-to-year operational tempo during the last decade and a half has affected demands for UV missions.

## Table A-1. MQ-1 Predator Mishap Statistics

| | CLASS A (ALL) | | | | | DESTROYED | | | | | CLASS B | | | | | CLASS A + CLASS B | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| YEAR | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | HOURS | C-HOURS |
| FY96 | 1 | 1 | 500.00 | 500.00 | | 1 | 1 | 500.00 | 500.00 | | 0 | 0 | 0.00 | 0.00 | | 1 | 1 | 500.00 | 500.00 | | 200 | 200 |
| FY97 | 3 | 4 | 122.10 | 150.55 | 6.95 | 3 | 4 | 122.10 | 150.55 | 6.10 | 0 | 0 | 0.00 | 0.00 | 1.45 | 3 | 4 | 122.10 | 150.55 | 8.40 | 2457 | 2657 |
| FY98 | 0 | 4 | 0.00 | 67.62 | 6.77 | 0 | 4 | 0.00 | 67.62 | 5.93 | 0 | 0 | 0.00 | 0.00 | 1.45 | 0 | 4 | 0.00 | 67.62 | 8.23 | 3258 | 5915 |
| FY99 | 2 | 6 | 38.95 | 54.30 | 6.79 | 2 | 6 | 38.95 | 54.30 | 5.94 | 0 | 0 | 0.00 | 0.00 | 1.45 | 2 | 6 | 38.95 | 54.30 | 8.24 | 5135 | 11050 |
| FY00 | 1 | 7 | 15.56 | 40.05 | 6.69 | 1 | 7 | 15.56 | 40.05 | 5.84 | 1 | 1 | 15.56 | 5.72 | 1.46 | 2 | 8 | 31.12 | 45.78 | 8.15 | 6426 | 17476 |
| FY01 | 4 | 11 | 52.83 | 43.92 | 6.65 | 4 | 11 | 52.83 | 43.92 | 5.80 | 1 | 2 | 13.21 | 7.98 | 1.40 | 5 | 13 | 66.04 | 51.90 | 8.06 | 7571 | 25047 |
| FY02 | 7 | 18 | 36.25 | 40.58 | 6.44 | 6 | 17 | 31.07 | 38.32 | 5.58 | 0 | 2 | 0.00 | 4.51 | 1.35 | 7 | 20 | 36.25 | 45.09 | 7.79 | 19313 | 44360 |
| FY03 | 2 | 20 | 9.75 | 30.83 | 6.08 | 2 | 19 | 9.75 | 29.29 | 5.27 | 0 | 2 | 0.00 | 3.08 | 1.37 | 2 | 22 | 9.75 | 33.92 | 7.45 | 20507 | 64867 |
| FY04 | 6 | 26 | 19.12 | 27.01 | 6.03 | 5 | 24 | 15.93 | 24.94 | 5.22 | 0 | 2 | 0.00 | 2.08 | 1.38 | 6 | 28 | 19.12 | 29.09 | 7.42 | 31383 | 96250 |
| FY05 | 10 | 36 | 24.38 | 26.22 | 5.77 | 9 | 33 | 21.94 | 24.04 | 5.00 | 1 | 3 | 2.44 | 2.19 | 1.41 | 11 | 39 | 26.81 | 28.41 | 7.18 | 41024 | 137274 |
| FY06 | 5 | 41 | 8.65 | 21.02 | 5.27 | 3 | 36 | 5.19 | 18.45 | 4.54 | 0 | 3 | 0.00 | 1.54 | 1.38 | 5 | 44 | 8.65 | 22.56 | 6.65 | 57798 | 195072 |
| FY07 | 7 | 48 | 8.84 | 17.50 | 5.13 | 5 | 41 | 6.31 | 14.95 | 4.52 | 0 | 3 | 0.00 | 1.09 | 1.44 | 7 | 51 | 8.84 | 18.60 | 6.57 | 79193 | 274265 |
| FY08 | 10 | 58 | 6.76 | 13.74 | 4.92 | 9 | 50 | 6.08 | 11.84 | 4.42 | 3 | 6 | 2.03 | 1.42 | 1.52 | 13 | 64 | 8.78 | 15.16 | 6.44 | 147980 | 422245 |
| FY09 | 13 | 71 | 6.94 | 11.65 | 4.70 | 10 | 60 | 5.34 | 9.84 | 4.22 | 4 | 10 | 2.13 | 1.64 | 1.46 | 17 | 81 | 9.07 | 13.29 | 6.16 | 187393 | 609638 |
| FY10 | 7 | 78 | 3.46 | 9.61 | 4.30 | 6 | 66 | 2.97 | 8.13 | 4.01 | 3 | 13 | 1.48 | 1.60 | 1.34 | 10 | 91 | 4.94 | 11.21 | 5.64 | 202330 | 811968 |
| FY11 | 12 | 90 | 5.01 | 8.56 | 4.50 | 11 | 77 | 4.60 | 7.32 | 4.27 | 5 | 18 | 2.09 | 1.71 | 1.30 | 17 | 108 | 7.10 | 10.27 | 5.81 | 239304 | 1051272 |
| FY12 | 9 | 99 | 4.18 | 7.81 | 4.30 | 9 | 86 | 4.18 | 6.79 | 4.14 | 3 | 21 | 1.39 | 1.66 | 0.99 | 12 | 120 | 5.57 | 9.47 | 5.29 | 215560 | 1266832 |
| FY13 | 9 | 108 | 4.53 | 7.37 | 4.37 | 9 | 95 | 4.53 | 6.48 | 4.11 | 1 | 22 | 0.50 | 1.50 | 0.77 | 10 | 130 | 5.03 | 8.87 | 5.14 | 198619 | 1465451 |
| FY14 | 8 | 116 | 4.20 | 7.01 | 4.20 | 7 | 102 | 3.68 | 6.16 | 3.68 | 2 | 24 | 1.05 | 1.45 | 1.05 | 10 | 140 | 5.25 | 8.46 | 5.25 | 190353 | 1655804 |

## Table A-2. MQ-9 Reaper Mishap Statistics

| | CLASS A (ALL) | | | | | DESTROYED | | | | | CLASS B | | | | | CLASS A + CLASS B | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| YEAR | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | HOURS | C-HOURS |
| FY01 | 0 | 0 | 0.00 | 0.00 | | 0 | 0 | 0.00 | 0.00 | | 0 | 0 | 0.00 | 0.00 | | 0 | 0 | 0.00 | 0.00 | | 30 | 30 |
| FY02 | 0 | 0 | 0.00 | 0.00 | 3.70 | 0 | 0 | 0.00 | 0.00 | 1.70 | 0 | 0 | 0.00 | 0.00 | 0.46 | 0 | 0 | 0.00 | 0.00 | 4.16 | 191 | 221 |
| FY03 | 0 | 0 | 0.00 | 0.00 | 3.70 | 0 | 0 | 0.00 | 0.00 | 1.70 | 0 | 0 | 0.00 | 0.00 | 0.46 | 0 | 0 | 0.00 | 0.00 | 4.17 | 100 | 321 |
| FY04 | 0 | 0 | 0.00 | 0.00 | 3.70 | 0 | 0 | 0.00 | 0.00 | 1.70 | 0 | 0 | 0.00 | 0.00 | 0.46 | 0 | 0 | 0.00 | 0.00 | 4.17 | 767 | 1088 |
| FY05 | 0 | 0 | 0.00 | 0.00 | 3.71 | 0 | 0 | 0.00 | 0.00 | 1.70 | 0 | 0 | 0.00 | 0.00 | 0.46 | 0 | 0 | 0.00 | 0.00 | 4.17 | 2373 | 3461 |
| FY06 | 2 | 2 | 62.89 | 30.12 | 3.72 | 0 | 0 | 0.00 | 0.00 | 1.71 | 0 | 0 | 0.00 | 0.00 | 0.47 | 2 | 2 | 62.89 | 30.12 | 4.19 | 3180 | 6641 |
| FY07 | 1 | 3 | 14.55 | 22.20 | 3.43 | 0 | 0 | 0.00 | 0.00 | 1.71 | 0 | 0 | 0.00 | 0.00 | 0.47 | 1 | 3 | 14.55 | 22.20 | 3.90 | 6872 | 13513 |
| FY08 | 3 | 6 | 22.24 | 22.22 | 3.31 | 0 | 0 | 0.00 | 0.00 | 1.73 | 0 | 0 | 0.00 | 0.00 | 0.47 | 3 | 6 | 22.24 | 22.22 | 3.78 | 13490 | 27003 |
| FY09 | 4 | 10 | 15.75 | 19.09 | 2.90 | 1 | 1 | 3.94 | 1.91 | 1.77 | 0 | 0 | 0.00 | 0.00 | 0.48 | 4 | 10 | 15.75 | 19.09 | 3.38 | 25391 | 52394 |
| FY10 | 1 | 11 | 1.78 | 10.14 | 2.35 | 1 | 2 | 1.78 | 1.84 | 1.68 | 0 | 0 | 0.00 | 0.00 | 0.50 | 1 | 11 | 1.78 | 10.14 | 2.85 | 56109 | 108503 |
| FY11 | 2 | 13 | 2.31 | 6.67 | 2.41 | 0 | 2 | 0.00 | 1.03 | 1.67 | 2 | 2 | 2.31 | 1.03 | 0.56 | 4 | 15 | 4.62 | 7.69 | 2.96 | 86526 | 195029 |
| FY12 | 4 | 17 | 3.39 | 5.43 | 2.43 | 3 | 5 | 2.54 | 1.60 | 1.99 | 1 | 3 | 0.85 | 0.96 | 0.22 | 5 | 20 | 4.24 | 6.39 | 2.65 | 118039 | 313068 |
| FY13 | 3 | 20 | 1.93 | 4.27 | 2.09 | 2 | 7 | 1.28 | 1.49 | 1.79 | 0 | 3 | 0.00 | 0.64 | 0.00 | 3 | 23 | 1.93 | 4.91 | 2.09 | 155793 | 468861 |
| FY14 | 4 | 24 | 2.23 | 3.70 | 2.23 | 4 | 11 | 2.23 | 1.70 | 2.23 | 0 | 3 | 0.00 | 0.46 | 0.00 | 4 | 27 | 2.23 | 4.16 | 2.23 | 179560 | 648421 |

## Table A-3. RQ-4 Global Hawk Mishap Statistics

| | CLASS A (ALL) | | | | | DESTROYED | | | | | CLASS B | | | | | CLASS A + CLASS B | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| YEAR | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | # | C-# | RATE | C-RATE | E-RATE | HOURS | C-HOURS |
| FY98 | 0 | 0 | 0.00 | 0.00 | | 0 | 0 | 0.00 | 0.00 | | 0 | 0 | 0.00 | 0.00 | | 0 | 0 | 0.00 | 0.00 | | 21 | 21 |
| FY99 | 1 | 1 | 383.14 | 354.61 | 5.53 | 1 | 1 | 383.14 | 354.61 | 3.68 | 0 | 0 | 0.00 | 0.00 | 1.84 | 1 | 1 | 383.14 | 354.61 | 7.37 | 261 | 282 |
| FY00 | 1 | 2 | 221.73 | 272.85 | 4.62 | 0 | 1 | 0.00 | 136.43 | 2.77 | 0 | 0 | 0.00 | 0.00 | 1.85 | 1 | 2 | 221.73 | 272.85 | 6.46 | 451 | 733 |
| FY01 | 0 | 2 | 0.00 | 164.07 | 3.71 | 0 | 1 | 0.00 | 82.03 | 2.78 | 0 | 0 | 0.00 | 0.00 | 1.85 | 0 | 2 | 0.00 | 164.07 | 5.56 | 486 | 1219 |
| FY02 | 2 | 4 | 127.71 | 143.63 | 3.73 | 2 | 3 | 127.71 | 107.72 | 2.79 | 0 | 0 | 0.00 | 0.00 | 1.86 | 2 | 4 | 127.71 | 143.63 | 5.59 | 1566 | 2785 |
| FY03 | 0 | 4 | 0.00 | 112.23 | 1.89 | 0 | 3 | 0.00 | 84.18 | 0.95 | 0 | 0 | 0.00 | 0.00 | 1.89 | 0 | 4 | 0.00 | 112.23 | 3.78 | 779 | 3564 |
| FY04 | 0 | 4 | 0.00 | 80.99 | 1.90 | 0 | 3 | 0.00 | 60.74 | 0.95 | 0 | 0 | 0.00 | 0.00 | 1.90 | 0 | 4 | 0.00 | 80.99 | 3.81 | 1375 | 4939 |
| FY05 | 0 | 4 | 0.00 | 51.41 | 1.93 | 0 | 3 | 0.00 | 38.56 | 0.96 | 1 | 1 | 35.20 | 12.85 | 1.93 | 1 | 5 | 35.20 | 64.27 | 3.86 | 2841 | 7780 |
| FY06 | 0 | 4 | 0.00 | 36.38 | 1.98 | 0 | 3 | 0.00 | 27.29 | 0.99 | 0 | 1 | 0.00 | 9.10 | 0.99 | 0 | 5 | 0.00 | 45.48 | 2.98 | 3214 | 10994 |
| FY07 | 0 | 4 | 0.00 | 24.06 | 2.05 | 0 | 3 | 0.00 | 18.05 | 1.02 | 0 | 1 | 0.00 | 6.02 | 1.02 | 0 | 5 | 0.00 | 30.08 | 3.07 | 5631 | 16625 |
| FY08 | 0 | 4 | 0.00 | 16.31 | 2.18 | 0 | 3 | 0.00 | 12.24 | 1.09 | 0 | 1 | 0.00 | 4.08 | 1.09 | 0 | 5 | 0.00 | 20.39 | 3.26 | 7894 | 24519 |
| FY09 | 1 | 5 | 13.75 | 15.73 | 2.38 | 0 | 3 | 0.00 | 9.44 | 1.19 | 0 | 1 | 0.00 | 3.15 | 1.19 | 1 | 6 | 13.75 | 18.87 | 3.57 | 7274 | 31793 |
| FY10 | 0 | 5 | 0.00 | 12.46 | 1.30 | 0 | 3 | 0.00 | 7.48 | 1.30 | 0 | 1 | 0.00 | 2.49 | 1.30 | 0 | 6 | 0.00 | 14.96 | 2.60 | 8322 | 40115 |
| FY11 | 1 | 6 | 7.56 | 11.25 | 1.46 | 1 | 4 | 7.56 | 7.50 | 1.46 | 1 | 2 | 7.56 | 3.75 | 1.46 | 2 | 8 | 15.11 | 15.00 | 2.92 | 13232 | 53347 |
| FY12 | 0 | 6 | 0.00 | 8.97 | 0.00 | 0 | 4 | 0.00 | 5.98 | 0.00 | 0 | 2 | 0.00 | 2.99 | 0.00 | 0 | 8 | 0.00 | 11.96 | 0.00 | 13520 | 66867 |
| FY13 | 0 | 6 | 0.00 | 7.11 | 0.00 | 0 | 4 | 0.00 | 4.74 | 0.00 | 0 | 2 | 0.00 | 2.37 | 0.00 | 0 | 8 | 0.00 | 9.48 | 0.00 | 17542 | 84409 |
| FY14 | 0 | 6 | 0.00 | 5.53 | 0.00 | 0 | 4 | 0.00 | 3.68 | 0.00 | 0 | 2 | 0.00 | 1.84 | 0.00 | 0 | 8 | 0.00 | 7.37 | 0.00 | 24164 | 108573 |

Table A-4. MQ-1 Predator Mishap Summary (A: 116 Total Mishaps, B: 24 Total Mishaps)

| | First 10 Years | Last 9 Years | First Half of Hours (15 Years) | Last Half of Hours (4 Years) | FY14 | Total Life |
|---|---|---|---|---|---|---|
| **Class A Rate** | 26.2 | 5.3 | 9.6 | 4.5 | 4.2 | 7.0 |
| **Class A+B Rate** | 28.4 | 6.7 | 11.2 | 5.8 | 5.3 | 8.5 |
| **Hours** | *137,274* | *1,518,530* | *811,968* | *843,836* | *190,353* | *1,655,804* |

Table A-5. MQ-9 Reaper Mishap Summary (A: 24 Total Mishaps, B: 3 Total Mishaps)

| | First 7 Years | Last 7 Years | First Half of Hours (12 Years) | Last Half of Hours (2 Years) | FY14 | Total Life |
|---|---|---|---|---|---|---|
| **Class A Rate** | 22.2 | 3.3 | 5.4 | 2.1 | 2.2 | 3.7 |
| **Class A+B Rate** | 22.2 | 3.8 | 6.4 | 2.1 | 2.2 | 4.2 |
| **Hours** | *13,513* | *634,908* | *313,068* | *335,353* | *179,560* | *648,421* |

Table A-6. RQ-4 Global Hawk Mishap Summary (A: 6 Total Mishaps, B: 2 Total Mishaps)

| | First 9 Years | Last 8 Years | First Half of Hours (14 Years) | Last Half of Hours (3 Years) | FY14 | Total Life |
|---|---|---|---|---|---|---|
| **Class A Rate** | 36.4 | 2.1 | 11.3 | 0.0 | 0.0 | 5.5 |
| **Class A+B Rate** | 45.5 | 3.1 | 15.0 | 0.0 | 0.0 | 7.4 |
| **Hours** | *10,994* | *97,579* | *53,347* | *55,226* | *24,164* | *108,573* |

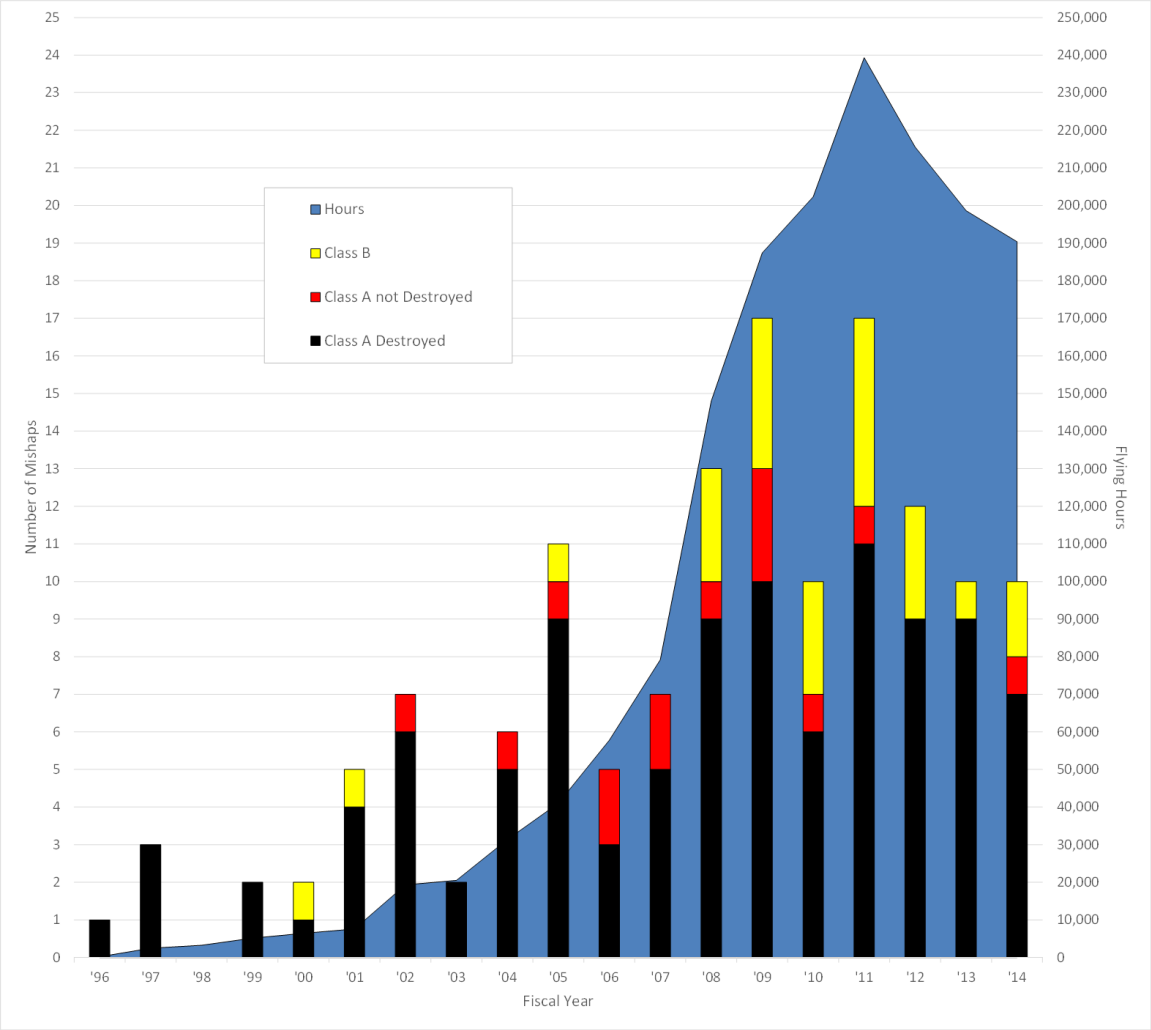



Figure A-1. MQ-1 Predator Mishaps and Flying Hours per Fiscal Year

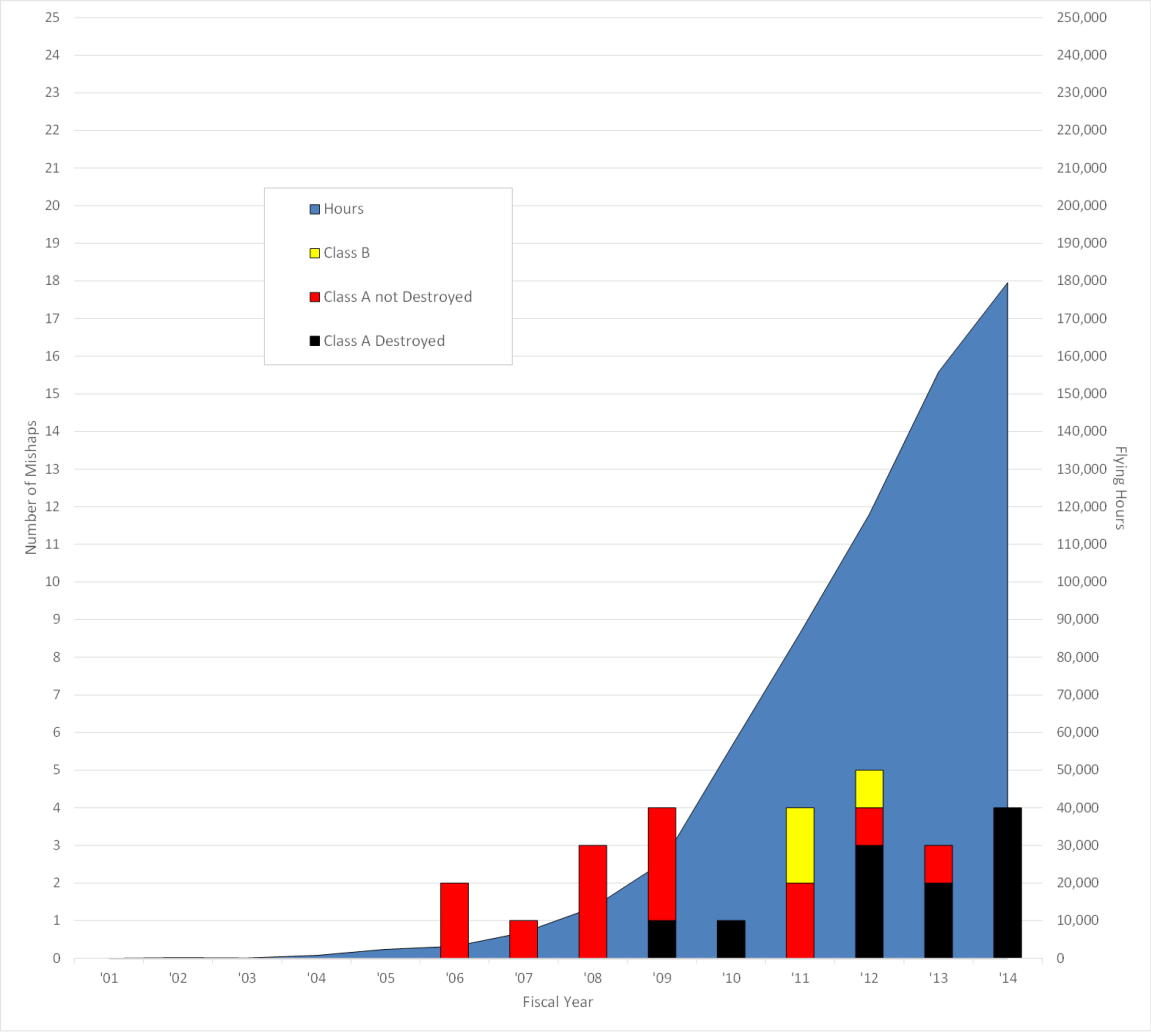



Figure A-2. MQ-9 Reaper Mishaps and Flying Hours per Fiscal Year

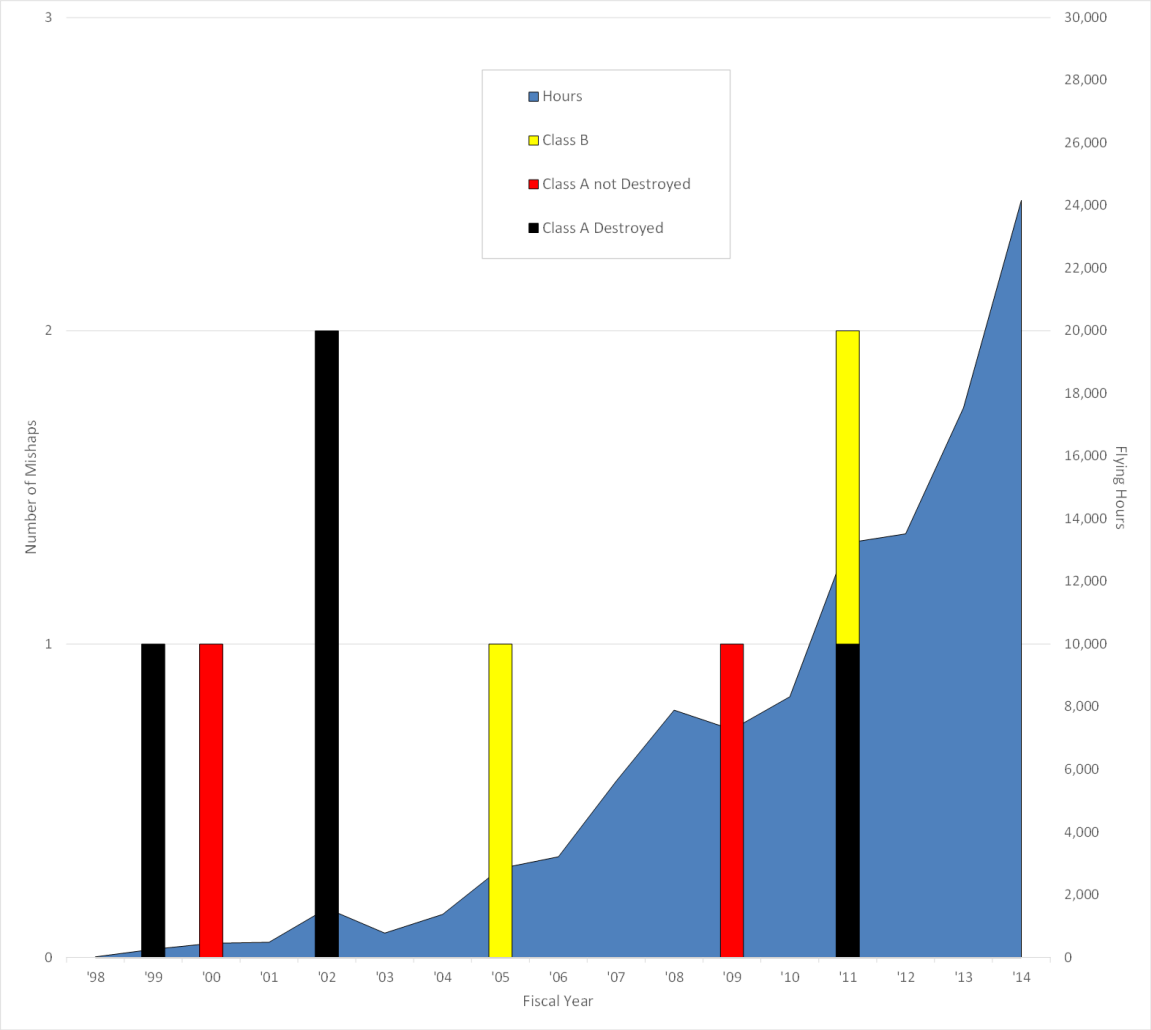



Figure A-3. RQ-4 Global Hawk Mishaps and Flying Hours per Fiscal Year

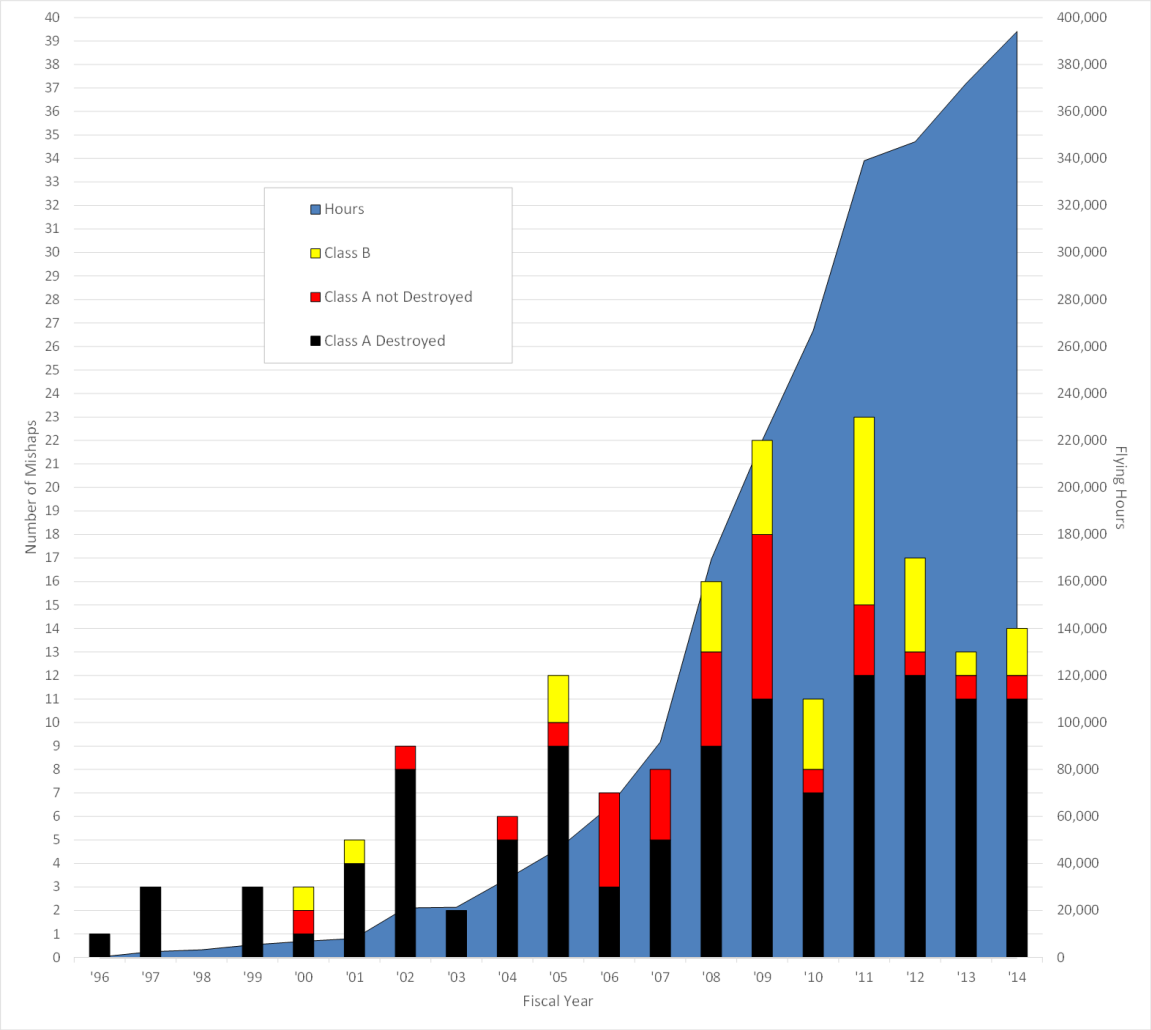



Figure A-4. Combined UV Mishaps and Flying Hours per Fiscal Year

## Accident Report Data

Tables A-7 through A-9 were constructed from the AF legal repository for accident investigation board (AIB) reports.[4] Accident boards may be publicly disclosed (unlike safety investigation reports) and are mandatory for any Class A mishap, per AFI 51-503 [290]. Of 134 Class A mishaps for the MQ-1, MQ-9, and RQ-4 through FY13, 82 were immediately accessible on the repository. Class B through E data were not sought, as those incidents are not investigated or reported on as thoroughly as Class A incidents, and it has been found in the past that not enough detail exists in their available write-ups with regard to software and human causal scenarios [293].

Each available report was analyzed for the board's opinion on the primary and contributing causal factors to the mishap. Reports from FY09 and earlier only had an executive summary available. Reports from FY10 and later were presented whole (albeit sans appendices). For those, only the executive summary was referenced to determine causal factors—although sometimes within the rest of the report, auxiliary factors were mentioned.[5] When an executive summary was not included in some of the complete reports, the investigator conclusion (which contains equivalent information) was queried.

Six categories of causal factor were searched for: Hardware, HF, software, loss of UV spatial awareness, lost/degraded data link, and pre-mission considerations. The first three (hardware, software, and HF) are mutually exclusive. Hardware refers to any mechanical or electrical defect or failure. Software refers to defective flight control laws, poorly designed control interfaces, or anomalies in the code. HF explicitly covers causal factors attributed to human error (e.g., procedural violations, loss of situation awareness, etc.).

The remaining three categories overlap with the first three. Loss of spatial awareness is a subset of HF describing incidents in which the human operator lost the ability to sense and avoid an unfavorable UV position or attitude with respect to the environment. Link quality issues (between the ground station and the UV) were cataloged when mentioned in the causal factor discussions; they could be caused by hardware or software, and sometimes the investigations could not determine the reason. Finally, pre-mission considerations mentioned in the reports included design issues, maintenance contributions, policy flaws, and culture (often in the form of organizational pressure to execute a high workload of missions).

For each type of causal factor, table A-7 shows the percentage of mishap reports (of each UV) that cite it, regardless of whether it was primary or contributing. The percentages do not sum to 100% because the categories are not all mutually exclusive and because some mishap reports cited multiple causal factors. Tables A-8 through A-13 contain the raw data from the report analyses, including a short summary of each. An “X” denotes a primary cause, while a “c” denotes a contributing factor.

---

[4] http://usaf.aib.law.af.mil/

[5] Human-Factors Analysis and Classification System (HFACS) discussions and error codes, for example.

Table A-7. Mishap Factor Contributions

| Type (# Mishaps) | Hardware | Human Factors | Loss of UV Spatial Awareness | Software | Lost or Degraded Link | Design / Pre Mission |
|---|---|---|---|---|---|---|
| All UVs (82) | 60% | 41% | 7% | 12% | 12% | 40% |
| Predator (74) | 62% | 41% | 8% | 12% | 14% | 38% |
| Reaper (5) | 20% | 80% | 0% | 0% | 0% | 60% |
| Global Hawk (3) | 67% | 0% | 0% | 33% | 0% | 67% |

The Predator contributed to most of the reported mishaps, and more than half of its mishaps had an element of hardware failure. Just under half of its mishaps involved HF. The Reaper, being a refined physical design to the Predator, was signified by a larger percentage—80%, which is the aviation industry average—of HF contributions. The Global Hawk, having the least number of mishaps, also showed high contributions from hardware and a contribution from software issues that was relatively higher than the other UVs. All UVs cited a sizable contribution from pre-mission considerations, while the contributions from poor link quality or loss of spatial sensing were not as high as might be believed from public sentiment.

It should be emphasized that even with teams of investigators working on accident analyses, results between reports will vary depending on the composition of the team investigating [4]. Different analysis methods will have varying assumptions as to which factors are the most important, with factors considered tangential receiving "superficial" analysis [21, p. 146]. It was also found recently that the correlation of HF findings between different groups analyzing the same accidents was only 60 percent [293]. Furthermore, even complete accident reports do not represent the full picture or complete opinion of practitioners. That information is contained in safety reports, which are privileged documents only available to organizational stakeholders. Thus, the data and percentages reported here—in addition to being useful only as descriptive statistics—are based on information that has been sanitized for public consumption.

Table A-8. Mishap Factor Raw Data (1 of 6)

| DATE | AIRCRAFT | LOCATION | Mech | Human | S&A | Software | Link | Pre Msn | |
|---|---|---|---|---|---|---|---|---|---|
| 1999-12-06 | RQ-4A | Edwards AFB, CA | | | | X | | c | Execution of excessive commanded ground speed introduced by a combination of of AFMSS and GH software problems, not recognized by mission planning or validation processes |
| 2000-09-14 | RQ-1L | Nellis AFB, NV | | X | | c | | c | Pilot activation of wrong menu during flight: result of habit developed over time and sense of being rushed. Contributing factor is basic design of the UV control system: menu systems allows placement of system in a hazardous condition without warning or verification |
| 2000-10-04 | RQ-1L | Grey Butte, CA | | | | X | | c | Execution of full tail up by the flight computer, caused by a failure in the flight computer. The watchdog timer had been disabled to reduce cost and schedule impacts of a software update |
| 2000-10-23 | RQ-1K | Kosovo | X | | | | | c | Failure of the propeller pitch system causing a reverse pitch. Maintenance actions were contributing |
| 2001-03-30 | RQ-1L | Bosnia | | X | | | | | Pilot's failure to immediately execute critical checklist steps for pitot static icing, after nonuse of the pitot heating system |
| 2001-12-30 | RQ-4A | AOR | X | | | | | c | Structural failure of the right V-tail and ruddervator due to massive delamination of main spar during turbulence. Improperly installed actuator nut plate bolt caused failure |
| 2002-01-22 | RQ-1L | AOR | | X | | | | | Crew's failure to complete checklist items in proper order |
| 2002-01-25 | RQ-1B | AOR | | X | | | | | Pilot's inability to maintain control during strong wind gusts |
| 2002-05-17 | RQ-1L | AOR | | | | | | X | Incorrect assembly of right tail plane control servo by manufacturer |
| 2002-05-25 | RQ-1B | Kuwait | X | | | | | | Catastrophic failure of the engine #3 cylinder rocker arm |
| 2002-07-10 | RQ-4A | AOR | X | | | | | | Fuel nozzle failed in the high flow condition eventually causing full failure of the engine |
| 2002-09-17 | RQ-1L | Southwest Asia | | X | c | | c | | Unintentionally flown into a hazardous cloud. Aircraft entered an area of meteorological activity, data link was lost for 20 sec, reestablished, lost again. |
| 2002-10-25 | RQ-1L | Nellis AFB, NV | | X | c | | c | | Inattention to altitude on the part of the aircrew (loss of SA, fixation on landing checklist, realizing terrain masking too late, and not initiating a lost link procedure on time). |
| 2003-01-01 | RQ-1B | AOR | X | | | | | c | Total loss of engine oil resulting in seized engine. Inattention by MX during engine inspection and improper technical procedures |

Table A-9. Mishap Factor Raw Data (2 of 6)

| DATE | AIRCRAFT | LOCATION | Mech | Human | S&A | Software | Link | Pre Msn | |
|---|---|---|---|---|---|---|---|---|---|
| 2003-12-11 | RQ-1 | AOR | | X | | c | c | | Abrupt pitch inputs made by pilot during a nose-high unusual attitude after disengaging autopilot. Crew icing analysis had led to decision to disengage autopilot. Software anomaly had caused nose-high attitude without pilot awareness. Intermittent link connectivity during ensuing pitch oscillations after disengaging AP |
| 2004-06-14 | MQ-1L | Nellis, NV | | X | | | | c | Late executed go-around from a poorly executed approach. Poor ORM and improper runway barrier placement contributed |
| 2004-08-17 | MQ-1L | AOR | X | | | | | c | Engine fire caused by non-standard routing of oil pump supply line. No clear MX guidance on how to route the hoses mixed with high ops tempo. |
| 2004-09-22 | MQ-1L | Nellis, NV | | X | c | | | c | Failure of pilot to correct a high flare in time to prevent a hard landing. Windshear during flare caused loss of airspeed. Did not initiate go around earlier when it was already a poor approach. Sensor operator did not provide calls of excessive airspeed and VIS during poor approach. Pilot tried a late go-around instead of reducing power after the impact. Poor sensory cues contributed to decision to initiate a go-around after impact. |
| 2004-10-13 | MQ-1B | Grey Butte CA, | | X | | | | c | Student caused PIO and IP failed to take appropriate corrective action. Poor training of students about PIO |
| 2004-11-24 | MQ-1L | AOR | | X | | | | c | Pilot error. Failed to execute landing checklist after troubleshooting a nav malfunction and deciding to land immediately, inapproprately leaving airspeed hold on. Training, flight discipline, and supervision contributed |
| 2005-01-14 | MQ-1L | AOR | | | | c | X | c | Loss of all control of the UV through all command links, eventual crash due to fuel starvation. Incorrect procedures, training, data corruption causing GCS reset, supervision, ops tempo. |
| 2005-03-27 | MQ-1L | AOR | X | | | | | c | Engine fire caused by fuel leak. No dedicated fire suppresison system, inadequate procedures for fires, inadequate MX tech data for routing fuel lines or preflight inspections |
| 2005-03-30 | MQ-1L | AOR | X | | | | | | Failure of bearing on propeller shaft |
| 2005-10-21 | MQ-9A | Grey Butte, CA | | X | | | | | Pilot failed to control approach and initiated go-around too late. Inadequate supervision, breakdown of CRM. |

Table A-10. Mishap Factor Raw Data (3 of 6)

| DATE | AIRCRAFT | LOCATION | Mech | Human | S&A | Software | Link | Pre Msn | |
|---|---|---|---|---|---|---|---|---|---|
| 2006-03-20 | MQ-1 | AOR | | X | | | | | Pilot used poor judgement and turned off SAS control under Ku-band control, resulting in aircraft banking excessively, losing link, and crashing. |
| 2006-06-22 | MQ-1L | Creech AFB, NV | X | | | | | c | Loose oil fiter caused loss of engine oil pressure. Lack of mechanical markings to ensure proper installation. |
| 2006-08-03 | MQ-1B | Creech AFB, NV | X | | | | | | Pilot error. Inadvertently shut off engine when trying to raise landing gear, and lack of airmanship by not attempting a restart |
| 2007-01-17 | MQ-1B | AOR | X | | | | | | Cracked crankshaft caused engine failure |
| 2007-02-23 | MQ-1B | AOR | X | | | | | | Failure of propeller pitch servo motor |
| 2007-03-26 | MQ-1B | Kandahar | | X | c | c | | c | Pilot error. Confused a runway bounce with a flare pitchup, did not command high enough pitch to prevent subsequent bounces. Lack of visual cues of attitude, and lack of pilot feedback and unique control logic contribute to create a situation in which aircrew is unable to recognize the proper control inputs for effective landing |
| 2007-07-30 | MQ-1B | Balad | X | | | | | | Failed ignition module |
| 2007-07-31 | MQ-1L | Balad | X | | | c | | c | Failure of manifold pressure sensor. No visual or aural warnings by design, and no checklist item to consider diagnosing the pressure sensure when the engine fails. |
| 2007-11-29 | MQ-1L | Ali | X | | | | | | Failure of right tail board receiver |
| 2007-12-17 | MQ-1B | AOR | X | | | | | | Short circuit of alternator |
| 2008-04-09 | MQ-1B | AOR | X | | | | | | Engine throttle body assembly causing engine to fail |
| 2008-05-02 | MQ-1B | Ali | X | | | | | | Failed ignition module |
| 2008-06-02 | MQ-1B | Balad | X | | | | | | Electrical system overload |
| 2008-06-12 | MQ-1B | Kandahar | X | | | | | | Faulty connection, no determination as to why |
| 2008-07-21 | MQ-1B | AOR | c | | | | X | | Power outage at GCS caused a lost link. UV was never found |
| 2008-08-01 | MQ-1B | Balad | X | | | | | | Short circuit of electrical system and alternator |
| 2008-10-19 | MQ-1B | Ali | X | | | | | c | Failure of bearing on propeller shaft. Use of old out of spec installation tool contributed. |
| 2008-11-02 | MQ-1B | Kandahar | | X | | | | c | Aggressive turn with an asymmetrically loaded aircraft. |
| 2009-02-22 | MQ-1B | AOR | X | | | | | | Flight computer electrical failure |
| 2009-03-20 | MQ-9 | Fort Irwin, CA | | | | | | X | Incorrectly assembled oilt temp valve |
| 2009-04-20 | MQ-1B | Jalalabad | X | | | | | | Catastrophic electrical failure |
| 2009-04-28 | MQ-1B | Creech AFB, NV | X | | | | | | Disconnected manifold pressure line |
| 2009-05-08 | MQ-1B | Kandahar | X | | | | | c | Partially dislodged control chip. Assymetric loading contributed |
| 2009-05-13 | MQ-1B | AOR | | | | | X | | Lost link, no clear cause determined |

Table A-11. Mishap Factor Raw Data (4 of 6)

| DATE | AIRCRAFT | LOCATION | Mech | Human | S&A | Software | Link | Pre Msn | |
|---|---|---|---|---|---|---|---|---|---|
| 2009-08-13 | MQ-1B | Balad | X | | | | | c | Failure of variable pitch propeller shaft, due to improper tempering at time of manufacture |
| 2009-09-04 | MQ-1B | Kandahar | X | c | | | | | Variable pitch servo malfunction. Aircrew did not realize they could have fired weapons to reduce aircraft weight. |
| 2009-09-14 | MQ-1B | Balad | X | | | | | | Failure of left tail servo |
| 2009-10-03 | MQ-1B | Kandahar | | X | c | | | | HFACS included. Pilot error. CFIT: Channelized attention (combat) away from flying the UV. |
| 2009-11-20 | MQ-1B | Kandahar | X | | | | | | Catastrophic electrical failure |
| 2010-04-20 | MQ-1B | March AFB, CA | | X | | c | | c | IP and student did not notice airspeed too low. Poor training and experience contributed, as well as deficient tng program, poor interface design |
| 2010-07-28 | MQ-1B | Cannon AFB | | X | | | c | c | HFACS included. Shadow pilot deviated from checklist procedures and turning her uplink on, hijacking the command signal. Contributing were a failure of having req crew of 2 in the shadow trailer, absence of published directives for shadow ops, failure to understand LOS issues of portable GCS which was acting as primary. |
| 2010-08-16 | MQ-1B | Balad | | X | | | | | HFACS included. Pilot error. Did not correctly execute procedure of turning stability system on beforte takeoff due to checklist error, inattention, and expectancy. |
| 2010-09-19 | MQ-1B | AOR | X | | | | | | Oil leak and engine failure |
| 2010-12-09 | MQ-1B | Kandahar | | X | | | | c | HFACS included. Pilot inadvertently deactivated stability system, caused a stall, tried to reactivate SAS out of T.O. parameters (said he was unaware of them), did not attempt full stick deflection (max available means) to raise the nose. Lack of general UV training on SAS contributed, and switches are locasted very close to each other. |
| 2011-01-03 | MQ-1B | Kandahar | X | | | | | | Failure of altenator bearing |
| 2011-01-14 | MQ-1B | Djibouti | X | | | | | | Engine failure |
| 2011-03-15 | MQ-1B | Horn of Africa | c | X | | c | | | HFACS included. Pilot did misdiagnosed a stuck throttle servo (had seen similar indications on previous sorties which were different issues) and failed to perform the appropriate checklist. Pilot did not shutoff engine after landing despite many emergency landing checklists requiring so, and pilot did not have the judgement to resort to such a measure despite end of runway markings clearly visible. Stuck throttle warning was not set off by existing parameters. |

Table A-12. Mishap Factor Raw Data (5 of 6)

| DATE | AIRCRAFT | LOCATION | Mech | Human | S&A | Software | Link | Pre Msn | |
|---|---|---|---|---|---|---|---|---|---|
| 2011-05-01 | MQ-1B | Jalalabad | X | | | | | | Failure of thrust bearing that caused a clogged oil filter |
| 2011-05-05 | MQ-1B | Kandahar | c | X | | | | | HFACS included. Pilot failed to execute a successful engine out recovery after cooling system failure |
| 2011-05-07 | MQ-1B | Djibouti | X | | | | | | Aielron servo failure |
| 2011-05-17 | MQ-1B | Djibouti | c | X | c | | | | HFACS included. Inability to detect runway at night with infra red (weather) or GPS altitude (erroneously indicating component) |
| 2011-06-05 | MQ-1B | Jalalabad | | | | | X | | Loss of comm due to lightning strike |
| 2011-07-10 | MQ-1B | Jalalabad | | | | | X | | Lost link, no clear cause determined |
| 2011-08-20 | MQ-1B | AOR | X | | | | | c | Failure of variable pitch propeller shaft, due to lax of MX guidance defining serviceable life |
| 2011-12-13 | MQ-9A | Seychelles | X | c | | | | c | HFACS included. Electrical short in engine control assembly, contributed to by poor MX and pilot poorly executed forced landing. |
| 2012-01-30 | MQ-1B | Kandahar | | c | | | | X | HFACS included. Engine inspection failed to detect damage to coolant lines. LRE aircrew accidentally left LOS transmitter on when they pointed the antenna, causing hostile takeover and loss of altitude which could have helped it be a nicer landing |
| 2012-02-14 | MQ-1B | AOR | c | X | | | | | HFACS included. LRE pilot's failure to adequately execute flight manual procedures (including load shedding) for dual alternator failure. Poor CRM during checklists, judgement, channelized attention on a gear warning indicator. |
| 2012-02-21 | MQ-1B | Djibouti | X | | | | | | Failure of turbocharger bearing |
| 2012-04-04 | MQ-9A | Seychelles | | X | | | | c | HFACS included. Pilot accidentally moved throttle to OFF, cutting fuel. While performing engine out landing, aircrew failed to confirm landing gear extended. Limited experience cited |
| 2012-04-14 | MQ-1B | Jalalabad | X | | | | | c | Engine failure due to unique power cable installed during redundancy upgrades. |
| 2012-07-24 | MQ-1B | Jalalabad | | X | | c | | c | HFACS included. Pilot did not use nose camera for takeoff as is standard practice. Ball had an uncommanded slew during takeoff roll. Conflicting policy on camera use contributed as well as lack of camera confirmation in takeoff checklist |
| 2012-08-22 | MQ-1B | AOR | c | X | | | | | HFACS included. LRE pilot's failure to adequately execute flight manual procedures (including load shedding) for dual alternator failure |

Table A-13. Mishap Factor Raw Data (6 of 6)

| DATE | AIRCRAFT | LOCATION | Mech | Human | S&A | Software | Link | Pre Msn | |
|---|---|---|---|---|---|---|---|---|---|
| 2012-09-18 | MQ-1B | AOR | | | | | X | | Lost link, no clear cause determined |
| 2012-10-26 | MQ-1B | Jalalabad | c | X | | | | c | HFACS included. Innecessary movements of propeller pitch by pilot following a propeller servo motor failure. Insufficient checklist guidance, reinforced by incorrect sim training contributed. |
| 2012-12-05 | MQ-9 | Nellis AFB, NV | | X | | | | | HFACS included. Throttle was imporperly configured before the mission due to lack of checklist execution, causing reverse thrust when MCE pilot took UV out of autopilot and a stall |
| 2013-03-02 | MQ-1B | Kandahar | X | | | | | | Failure of ruddervator control chip |
| 2013-05-13 | MQ-1B | Creech AFB, NV | X | c | | | | | Variable pitch control cable experienced inconsistent current, causing a freeze of pitch angle for several hours then an abrupt thrust deficient angle. Pilot did not notice frozen pitch during ops checks. |
| 2013-06-27 | MQ-1B | Jalalabad | X | c | | | | c | HFACS included. Failed turbochrager forced emergency landing. LRE pilot initiated go-around with gusty winds, thinking he had enough thrust when he didn't, due to insufficient T.O. guidance |

# Appendix B

# Explicit Influence Map

The Explicit-Influence Map (EIM) is a conceptual data product I developed to support socio-organizational research sponsored by Draper Laboratory. It is currently assembled only manually using a freeware concept-mapping tool called Visual Understanding Environment (VUE).[1] Future work on the concept would include the capability to automatically produce the diagram using raw inputs in the form of policy documents and corresponding metadata that describes their relationships. However, the current format still offers some capabilities that are novel for a policy database.

These large charts are difficult to reproduce in document format. An example EIM is presented here as five separate figures (B-1 through B-7) as best as space allows. The figures show the entire EIM and each of its four quadrants to allow for a better resolution and legibility. Currently the optimal method for viewing an EIM is on a computer as a high-resolution, interactive portable document format (PDF). The original model can also be opened with VUE which allows more interaction when reading it as well as full editing privileges.

## Background

EIMs are intended to be used as common planning tools, similar to a STAMP safety-control structure. These maps in essence produce a visual representation of the explicit influences outlined in Figure 3-6 in Chapter 3 for a specific sociotechnical enterprise. The EIM traces all the explicitly documented influences from higher levels of an organization (or government) down through the lower levels of an enterprise and finally to local unit standards, techniques, and documented processes.

The map demonstrated here was built for the Air Force (AF) 412$^{th}$ Test Wing (412TW) and maps down to the functions of a flight-test squadron to support discussions in Chapter 4. Not just higher-level regulations and instructions are mapped, but local group and unit documents as well. A different AF enterprise in another location might have an EIM that looks similar to this one at higher levels, but very different in the lower levels. Although I spent several months compiling the documents for the example EIM, particularly at the lower levels, it is the expert practitioners who should construct and

[1] Tufts University, © 2013. http://vue.tufts.edu/

maintain the actual product. It is likely that the example EIM is not complete because I was only able to consult with local practitioners occasionally to make updates.

I noted during my research of AF regulations and instructions that policy documents were inconsistently formatted, updated, referenced, and filed. Most of the regulations—alphanumerically searchable in a publicly-accessible online database[2]—contained on average one to three pages of front matter that discussed the history of revisions to the document, general references to other documents, and sometimes lengthier additional notes or memorandums giving justification for the current revision. Some documents instead contained references in an appendix; some directly discussed the higher regulation that was the parent document, and some did not. This is a filing and traceability structure that has not changed since before personal computers were invented. It is up to any casual reader to examine many of the documents several times to map their relationships and the flow of their authority.[3] The EIM is a method to instantaneously display all of the policy documents that affect the day-to-day operations of an enterprise (and, if desired, a unit or subsection within the enterprise). With no need for searching through reference lists or guessing search terms in an online database, the EIM visually represents all the documents with their authority relationships, and they can be fetched and read by simply selecting them on the map.

At face value, this visual format for organizing explicit documents is useful enough for merely indexing purposes. It is a living database that can easily and consistently be updated when documents are created, updated, split, merged, or retired. Appropriate use of color and shape coding would make it simple for practitioners to be notified of changes or updates. Instead of the information about references, justifications for changes, and document history being included within a document, that information would be included separately in its metadata for optional viewing in the database. This way, the formatting for the documents could be kept more consistent and with less boilerplate information in the documents themselves. In the current manual-VUE format, only the publication date has been added to the metadata, as well as a notes field that summarizes the key messages within some of the documents. Selecting one of the document blocks will automatically open a PDF of the actual document (or in some cases, a website).

Beyond the referencing abilities, the EIM encourages practitioners to be aware of redundancies, conflicts, gaps, and inconsistencies in their policies. This is important not just for the purpose of increasing efficiency of the organization, but because conflicts in policy can manifest as faulty control algorithms during the operating process as discussed in Chapter 3. The ability for all members of the organization to see and interact with the EIM encourages active employee involvement in policy feedback. If the EIM as an official product were to be managed as a database with consistent metadata, warnings could be

---

[2] http://www.e-publishing.af.mil/

[3] While 412TW practitioners are more than casual readers, I determined during fact-finding interviews that none of the personnel spoken with were aware of *all* the instructions that affect their work; this is no fault of their own, but the consequence of not having an easily graspable database of explicit influences. Even the practitioners in charge of maintaining a local repository of AF standards did not have copies of everything on the EIM.

issued when documents are changed or deleted which other documents reference. If end users have notes, tags, and annotations on these documents, they would be alerted that those documents have been affected. Another potential concept would be for written references made *within* the internal document to be tied to the metadata that links those documents to others, allowing for warnings to be given to authors to revise the language inside the documents when relationships change.

## Walkthrough

Figure 3-6 in Chapter 3 introduces four types of influences that might be found within a professional organization: organizational culture, behavioral standards, rules and techniques, and settings/configurations. An organization also has deliberate levels of authority. For example the AF has major commands, numbered air forces, wings, groups, and squadrons. Sometimes different names are used for these levels depending on local circumstances, but the levels of authority are very well defined in their respective realms. Above the AF exist organizational levels of the defense department, and so on. The types of influences found in an explicit document are not necessarily correlated to the organizational level from which the document originates. Culture and standards can be expressed all the way down at the squadron level, and vice versa, rules and techniques can be defined and standardized at higher levels. Even settings and configurations can come from higher levels. An example would be a temporary order from a major command designating a particular set of flight restrictions for the following week of operations.

Each block in the example EIM is color coded to represent the type of document that it refers to (e.g., law, directive, instruction, etc.). Other color-code strategies for the blocks could be employed to represent perhaps the level of organization, or the type(s) of influences found in the respective document. A simple graphical-interface filter could be used in a future iteration of the EIM to provide these as selectable options, but for this example the colors represent document types. *Black* represents national coded law. *Purple* represents government directives at all levels, while *green* represents instructions and regulations at all levels. *Magenta* ovals contain placeholders for some offices. The placeholders were inserted my discretion and are useful to illustrate policy flow even when intermediate policies cannot be found, are ambiguous, or do not exist.[4] *Orange* represents standards, handbooks, and technical orders, while *blue* represents guidance and academics. If a block is tinted in a darker shade, it is because it exists but is not publicly obtainable.

Arrows show the authority relationships between documents, with higher documents pointing to derived documents, supplements, and lower-level clarifications. When documents reside in the same or a similar level of organization, they share the same vertical position on the EIM. Horizontally, groupings of documents become naturally separated by themes, such as safety on the left, operations in the center, and acquisition on the right. Dashed lines show how guidance and standards complement directives and instructions. Horizontal lines between instructions and directives show where those

[4] These entities would likely not exist in future iterations. The EIM is not an organization chart. However, there may be benefit in allowing practitioners to insert notes and additional illustrations at the end-user level.

documents either cross-reference each other or where they are intended to be implemented in a complementary fashion. Two relationships are labeled near the top of the EIM. The Defense Test and Training Steering Group (DTTSG) is a working group specifically defined by the two policies surrounding it on the map. Similarly, the Defense Safety Oversight Council (DSOC) is defined and staffed per the policies surrounding it on the map. These tags were inserted at my discretion. In a future iteration of the EIM concept, tags and notes could similarly be entered by practitioners at the user level.

The power of a tool like the EIM becomes evident lower in the map. In the bottom center region is a collection of unit-level influences colored in *fuchsia* for the purpose of highlighting it for discussion. The lower that components are in the safety-control structure, the shorter the time constant and the faster that local information and activities tend to function. This also means that explicit influences from lower components tend to update and vary at a faster rate. This lower portion of the EIM might look different if it were constructed for a squadron managing the airfield instead of a squadron performing flight testing. Some higher-level influences might also look different depending on the lower office the EIM supports. Regardless, the EIM is capable of displaying explicit influences from all organizational levels as well as their relationships, and as policies change and become updated practitioners can use the EIM to immediately check for conflicts and comprehend the flow of authority. If an explicit influence exists, it will affect a controller during the operating process in some manner, either directly or through the lower influences it has authority over. All explicit influences belong on the EIM.

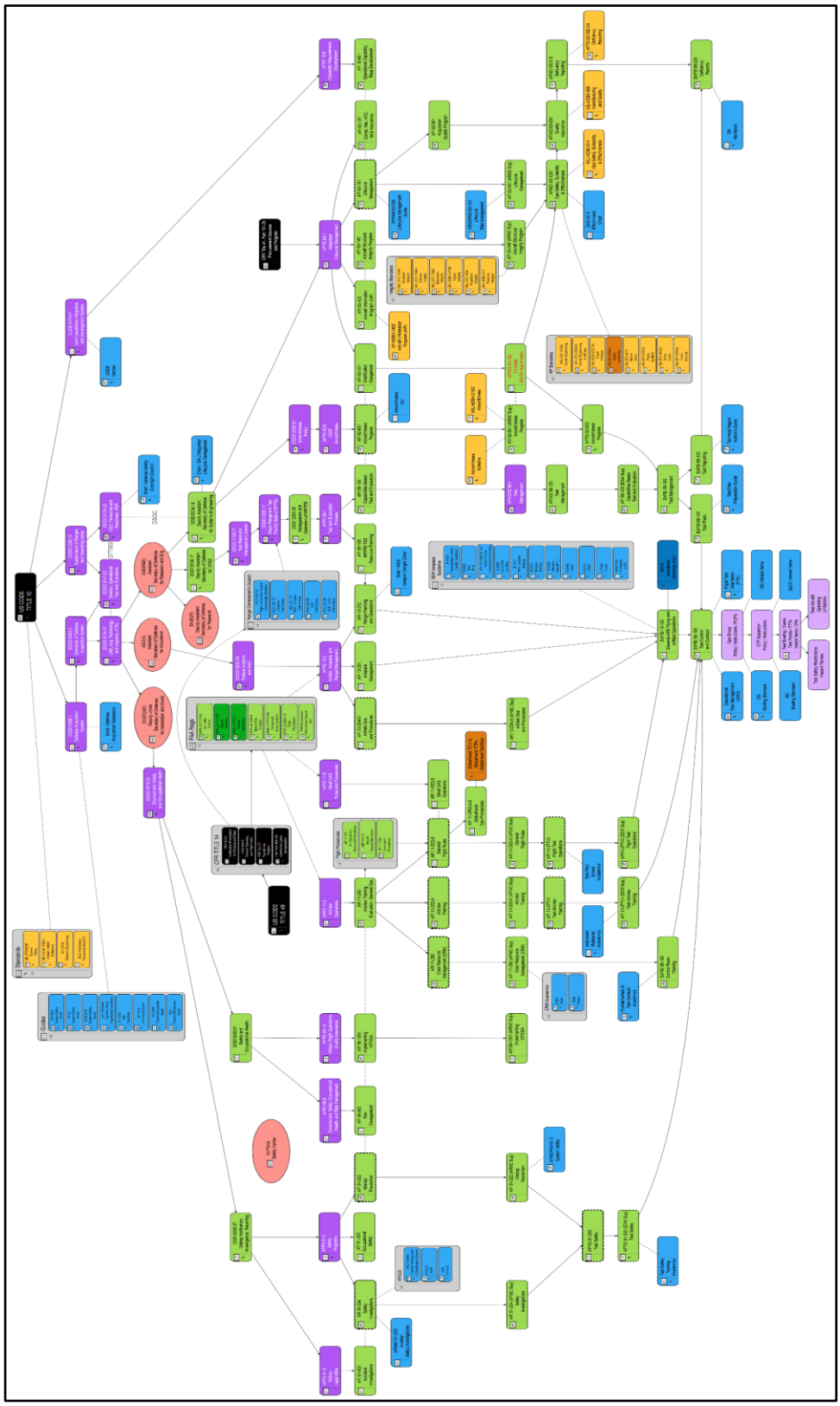


Figure B-1. Air Force Developmental Test Influence Map (Entire Diagram)

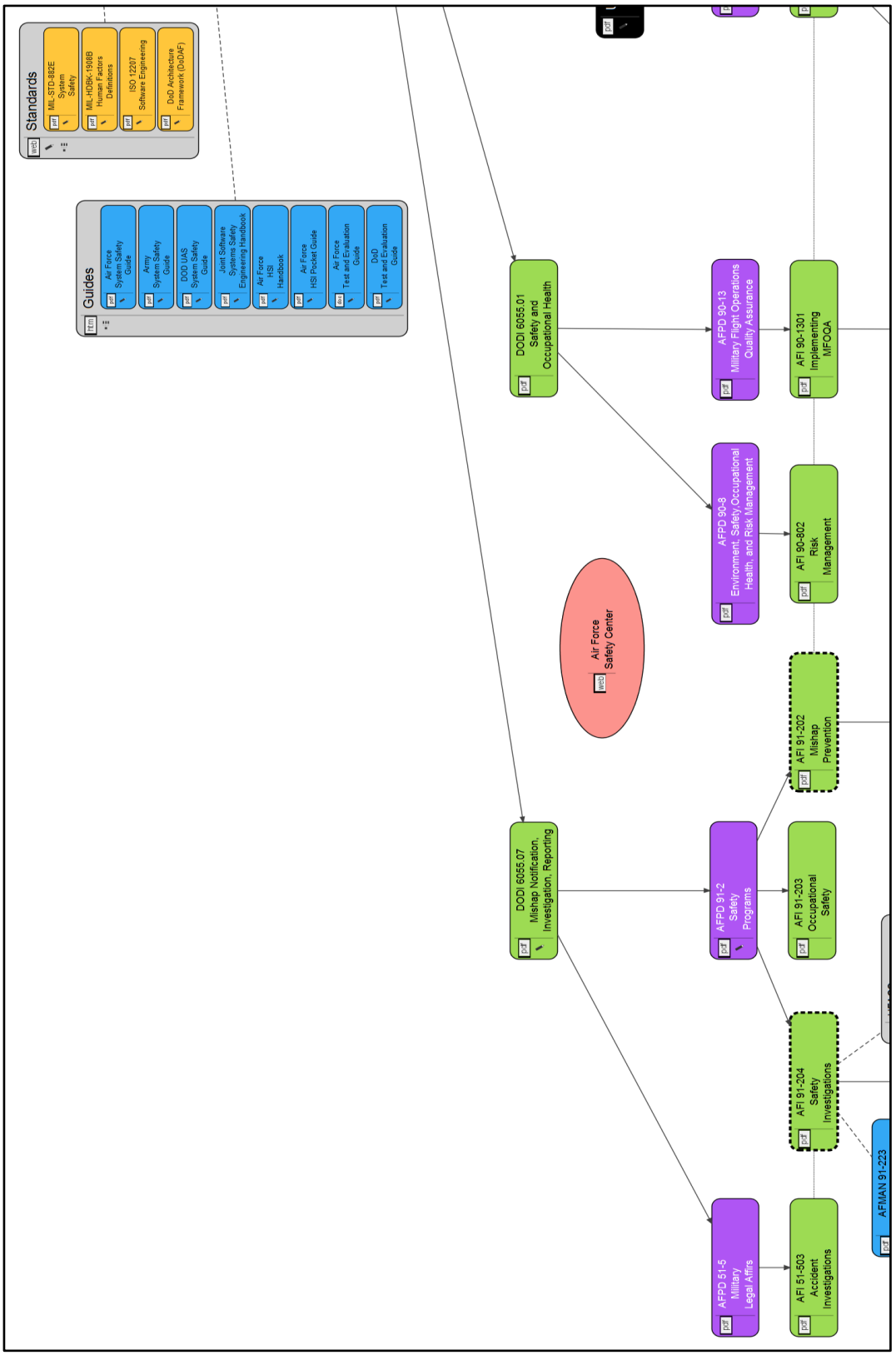



Figure B-2. Air Force Developmental Test Influence Map (Top Left)

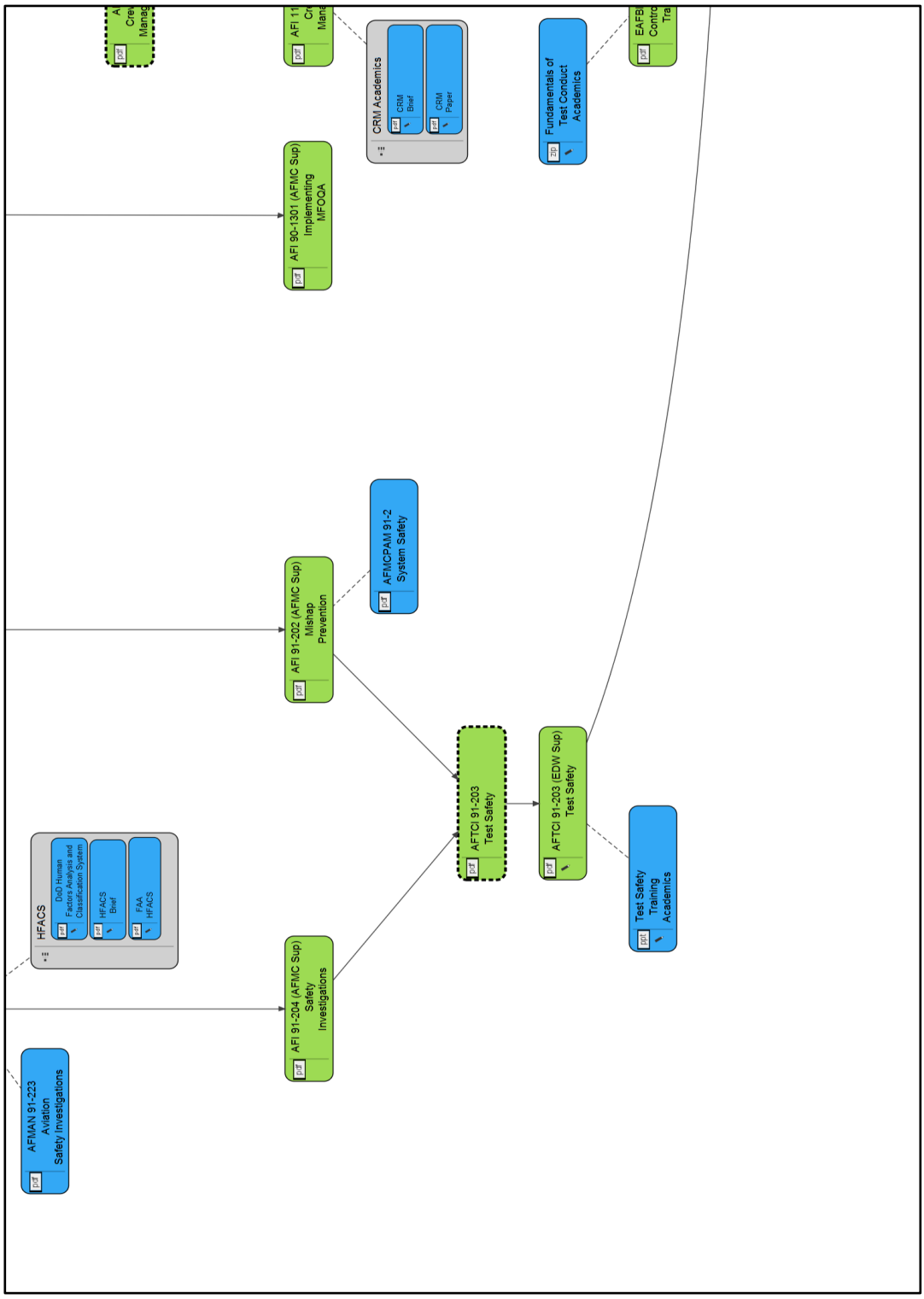



Figure B-3. Air Force Developmental Test Influence Map (Bottom Left)

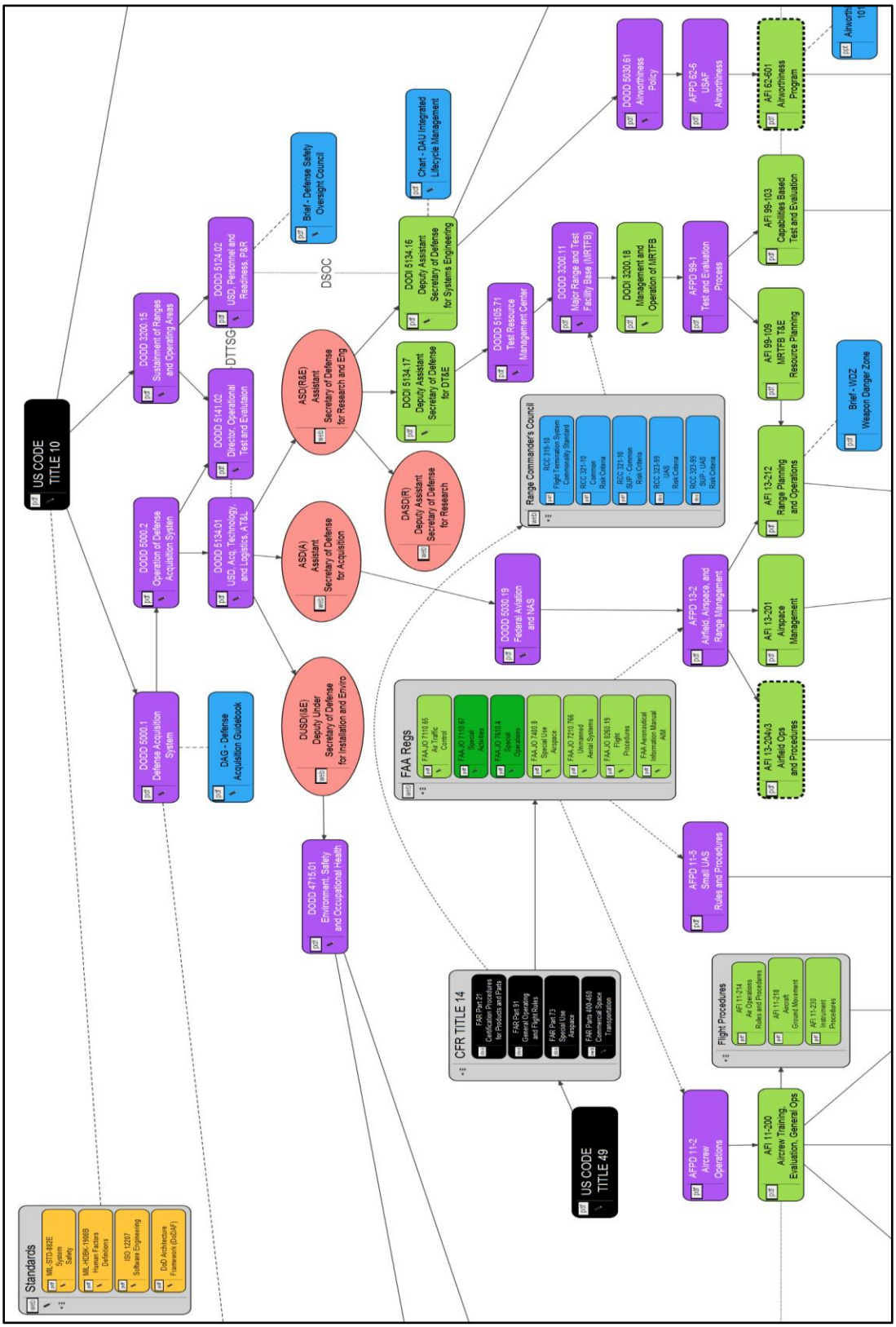



Figure B-4. Air Force Developmental Test Influence Map (Top Center)

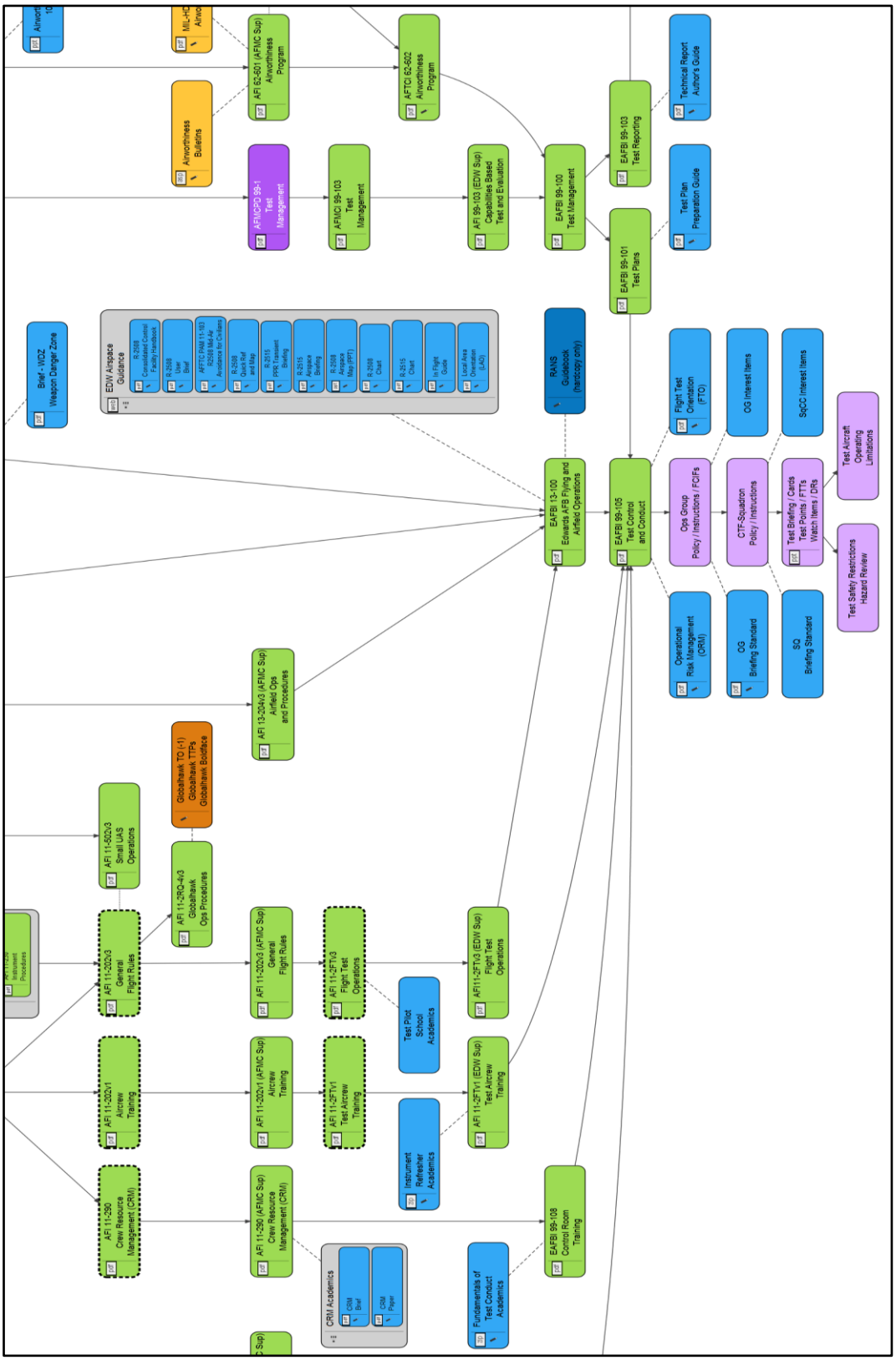



Figure B-5. Air Force Developmental Test Influence Map (Bottom Center)

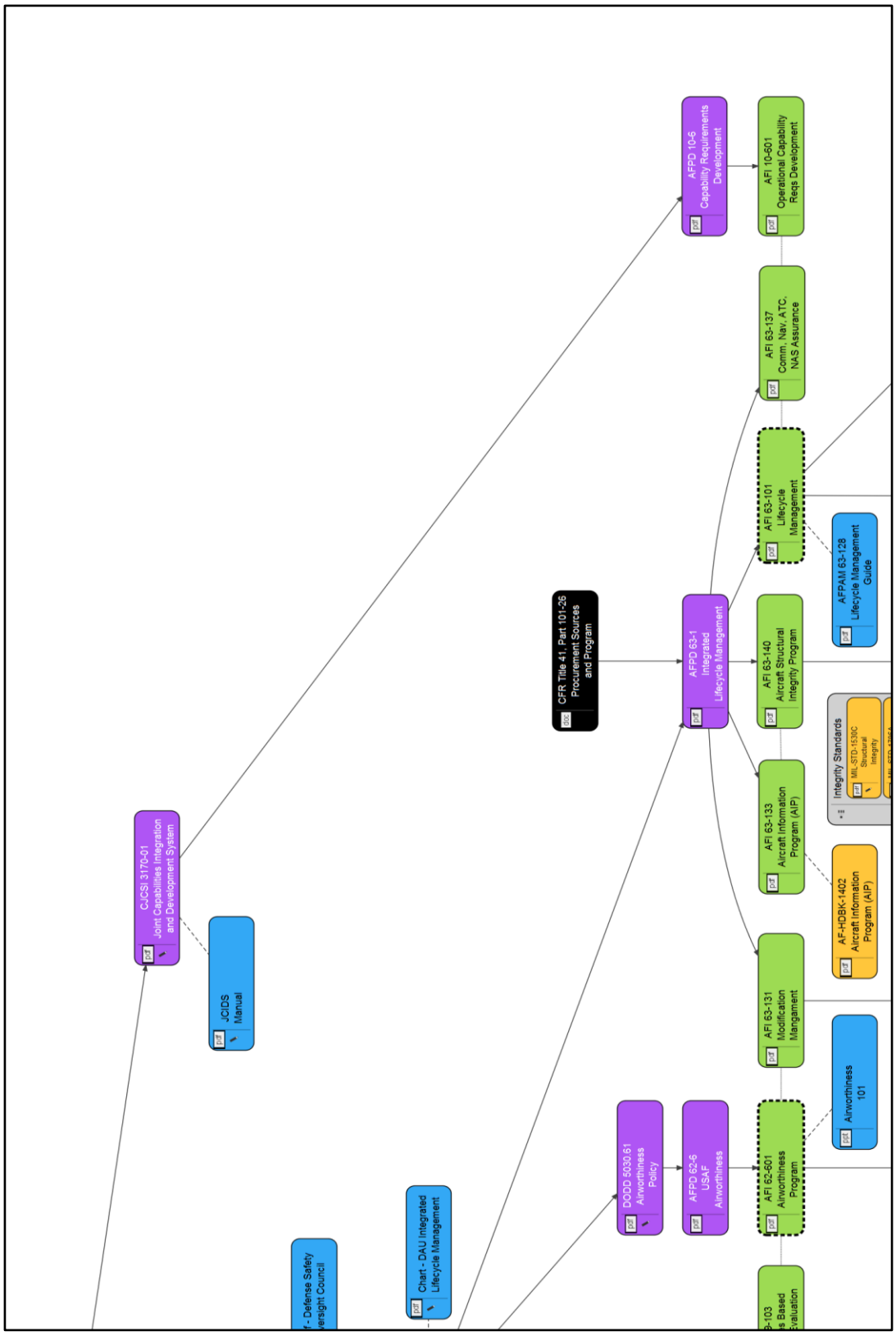



Figure B-6. Air Force Developmental Test Influence Map (Top Right)

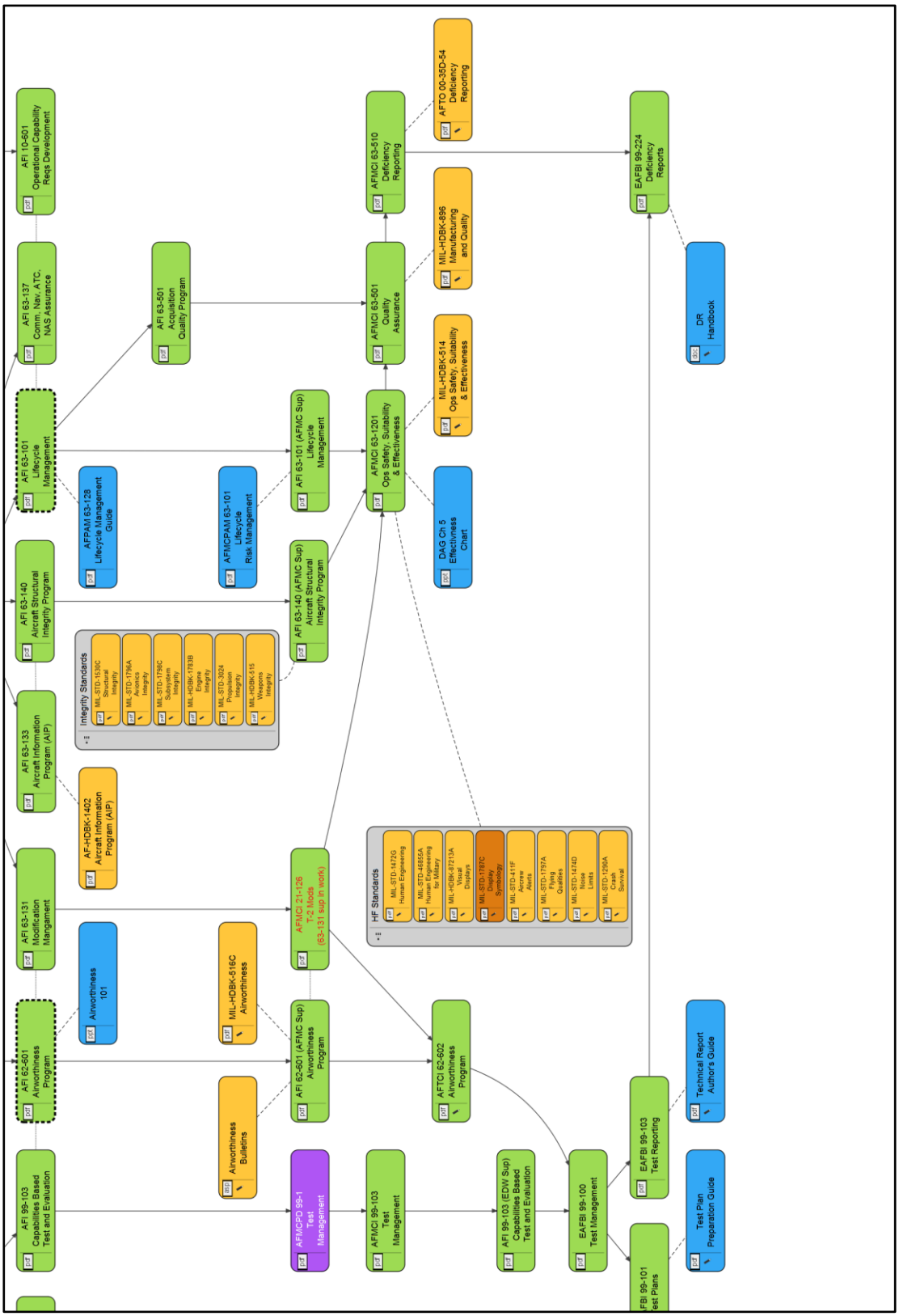



Figure B-7. Air Force Developmental Test Influence Map (Bottom Right)

# Appendix C

# Survey Data

The following information supplements the survey discussion in Chapter 4 by presenting raw data from the questionnaire, tabulated values, and the methodology and outcomes of the non-parametric statistical analyses (chi-square) performed on the data. The results are discussed in the chapter.

## Raw Multiple-Choice Data

Table C-1 contains the multiple-choice questions from the survey with corresponding reference codes. The three fields in blue (Intel-0, Inform-0, and Implem-0) were forced-choice questions. These asked participants to make a choice between the traditional or STPA plan for each of the three assessment types. The green fields were detailed questions. Those addressed specific attributes of each assessment type. Intelligibility had seven detailed questions; informativeness had six, and implementability had seven. There were thus twenty total detailed question. These asked participants to make a selection between "traditional plan", "STPA plan", "both equivalent", or "neither effective".

Table C-2 contains the raw responses from the eight participants. All survey plots and calculations presented in this thesis are built using this raw multiple-choice data. The columns of the table have been arranged so that participants are ordered into groupings by test-experience level, based on their responses to the demographic portion of the survey.

Table C-3 tabulates the proportion of preferences for each detailed-question field. These proportions are presented graphically in Chapter 4. For example, out of the eight total participants, Intel-1 received five choices for STPA, one choice for traditional, and two choices for both, resulting in the proportions shown in the first four entries of its corresponding row. Each of the twenty detailed questions had eight possible counts distributed among the four possible response choices.

Moving down the table, at the end of the detailed questions within each assessment type, a row is included that tabulates the average proportion of preferences for all detailed questions in that assessment type. Moving from left to right in the table, proportions are also calculated for two sub-samples. One is for people with one year or less of test experience (three participants), and the other is for people with more than one year of experience (five participants).

Table C-1. Multiple Choice Question Glossary

| Code | Question |
|---|---|
| **Intel-0** | **Which of the Safety Plans did you find MOST Intelligible?** |
| Intel-1 | Easy to quickly reference desired information |
| Intel-2 | Easy to read and comprehend |
| Intel-3 | Easy to find the "bottom line" |
| Intel-4 | Consistency of formatting across multiple similar entries (e.g., hazardous behaviors) |
| Intel-5 | Easier to mentally visualize the system |
| Intel-6 | Easy to understand what portions of the system are upgraded / being evaluated |
| Intel-7 | Easy to understand which equipment and personnel are part of only the testing (but not the intended field use) |
| **Inform-0** | **Which of the Safety Plans did you find MOST Informative?** |
| Inform-1 | Informative presentation of hazards (and unsafe actions, if applicable) |
| Inform-2 | Informative presentation of causes / causal scenarios |
| Inform-3 | Informative presentation of minimizing procedures / considerations |
| Inform-4 | Informative presentation of corrective actions |
| Inform-5 | Traceability of causes / causal scenarios to hazards / behaviors |
| Inform-6 | Traceability of minimizing procedures / considerations to causes / causal scenarios |
| **Implem-0** | **Which of the Safety Plans would you consider the MOST Implementable?** |
| Implem-1 | Ease of performing the hazard analysis |
| Implem-2 | Ease of constructing the safety plan document |
| Implem-3 | Ability for the format and information in the document to be used as a template for future documents |
| Implem-4 | Easy to teach the method to someone |
| Implem-5 | Perceived ability of analysis outputs to inform risk mitigation activities during test planning |
| Implem-6 | Perceived ability of analysis outputs to aid pre-mission briefs |
| Implem-7 | Perceived ability to implement changes to the safety planning as lessons are learned during test activites |

Table C-2. Multiple Choice Responses by Participant

| Experience ==> | 1 Year or Less | | | 2–5 Years | 5–10 Years | | | |
|---|---|---|---|---|---|---|---|---|
| Subject ==> | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| **Intel-0** | **STPA** | **STPA** | **STPA** | **STPA** | **STPA** | **STPA** | **Traditional** | **Traditional** |
| Intel-1 | STPA | Both | Both | STPA | STPA | STPA | STPA | Traditional |
| Intel-2 | Both | STPA | Both | STPA | STPA | STPA | Traditional | Traditional |
| Intel-3 | Both | STPA | STPA | STPA | STPA | STPA | Traditional | Traditional |
| Intel-4 | STPA | STPA | STPA | Both | STPA | STPA | STPA | STPA |
| Intel-5 | Both | STPA | Both | STPA | STPA | STPA | Traditional | STPA |
| Intel-6 | Both | STPA | STPA | Both | STPA | STPA | STPA | STPA |
| Intel-7 | Both | STPA | STPA | Both | STPA | STPA | STPA | STPA |
| **Inform-0** | **STPA** | **STPA** | **STPA** | **STPA** | **STPA** | **STPA** | **STPA** | **Traditional** |
| Inform-1 | STPA | STPA | STPA | STPA | STPA | STPA | STPA | STPA |
| Inform-2 | STPA | STPA | STPA | STPA | STPA | STPA | STPA | STPA |
| Inform-3 | STPA | STPA | Both | STPA | STPA | Traditional | STPA | Traditional |
| Inform-4 | STPA | STPA | Both | Both | STPA | Traditional | STPA | Traditional |
| Inform-5 | STPA | STPA | STPA | STPA | STPA | STPA | STPA | STPA |
| Inform-6 | STPA | STPA | STPA | STPA | STPA | Traditional | STPA | STPA |
| **Implem-0** | **STPA** | **STPA** | **STPA** | **STPA** | **STPA** | **STPA** | **STPA** | **Traditional** |
| Implem-1 | STPA | Both | STPA | STPA | Traditional | STPA | Both | Traditional |
| Implem-2 | STPA | Neither | STPA | Traditional | Traditional | STPA | Both | STPA |
| Implem-3 | STPA | STPA | STPA | Traditional | Traditional | Traditional | STPA | Traditional |
| Implem-4 | STPA | STPA | STPA | Traditional | Traditional | STPA | STPA | Traditional |
| Implem-5 | STPA | STPA | STPA | STPA | STPA | STPA | Both | STPA |
| Implem-6 | STPA | Neither | STPA | Traditional | STPA | STPA | STPA | Traditional |
| Implem-7 | STPA | Both | STPA | STPA | STPA | Both | Both | STPA |

Table C-3. Response Proportions, Detailed Questions

| | Proportion (All Respondents) | | | | Proportion (< 1 Year Experience) | | | | Proportion (> 1 Year Experience) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Detailed Question | STPA | Traditional | Both | Neither | STPA | Traditional | Both | Neither | STPA | Traditional | Both | Neither |
| Intel-1 | 0.625 | 0.125 | 0.250 | 0 | 0.333 | 0 | 0.667 | 0 | 0.800 | 0.200 | 0 | 0 |
| Intel-2 | 0.500 | 0.250 | 0.250 | 0 | 0.333 | 0 | 0.667 | 0 | 0.600 | 0.400 | 0 | 0 |
| Intel-3 | 0.625 | 0.250 | 0.125 | 0 | 0.667 | 0 | 0.333 | 0 | 0.600 | 0.400 | 0 | 0 |
| Intel-4 | 0.875 | 0 | 0.125 | 0 | 1 | 0 | 0.000 | 0 | 0.800 | 0 | 0.200 | 0 |
| Intel-5 | 0.625 | 0.125 | 0.250 | 0 | 0.333 | 0 | 0.667 | 0 | 0.800 | 0.200 | 0 | 0 |
| Intel-6 | 0.750 | 0 | 0.250 | 0 | 0.667 | 0 | 0.333 | 0 | 0.800 | 0 | 0.200 | 0 |
| Intel-7 | 0.750 | 0 | 0.250 | 0 | 0.667 | 0 | 0.333 | 0 | 0.800 | 0 | 0.200 | 0 |
| Intel-Avg | 0.679 | 0.107 | 0.214 | 0 | 0.571 | 0 | 0.429 | 0 | 0.743 | 0.171 | 0.086 | 0 |
| Inform-1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Inform-2 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Inform-3 | 0.625 | 0.250 | 0.125 | 0 | 0.667 | 0 | 0.333 | 0 | 0.600 | 0.400 | 0 | 0 |
| Inform-4 | 0.500 | 0.250 | 0.250 | 0 | 0.667 | 0 | 0.333 | 0 | 0.400 | 0.400 | 0.200 | 0 |
| Inform-5 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Inform-6 | 0.875 | 0.125 | 0 | 0 | 1 | 0 | 0 | 0 | 0.800 | 0.200 | 0 | 0 |
| Inform-Avg | 0.833 | 0.104 | 0.063 | 0 | 0.889 | 0 | 0.111 | 0 | 0.800 | 0.167 | 0.033 | 0 |
| Implem-1 | 0.500 | 0.250 | 0.250 | 0 | 0.667 | 0 | 0.333 | 0 | 0.400 | 0.400 | 0.200 | 0 |
| Implem-2 | 0.500 | 0.250 | 0.125 | 0.125 | 0.667 | 0 | 0 | 0.333 | 0.400 | 0.400 | 0.200 | 0 |
| Implem-3 | 0.500 | 0.500 | 0 | 0 | 1 | 0 | 0 | 0 | 0.200 | 0.800 | 0 | 0 |
| Implem-4 | 0.625 | 0.375 | 0 | 0 | 1 | 0 | 0 | 0 | 0.400 | 0.600 | 0 | 0 |
| Implem-5 | 0.875 | 0.000 | 0.125 | 0 | 1 | 0 | 0 | 0 | 0.800 | 0 | 0.200 | 0 |
| Implem-6 | 0.625 | 0.250 | 0 | 0.125 | 0.667 | 0 | 0 | 0.333 | 0.600 | 0.400 | 0 | 0 |
| Implem-7 | 0.625 | 0 | 0.375 | 0 | 0.667 | 0 | 0.333 | 0 | 0.600 | 0 | 0.400 | 0 |
| Implem-Avg | 0.607 | 0.232 | 0.125 | 0.036 | 0.810 | 0 | 0.095 | 0.095 | 0.486 | 0.371 | 0.143 | 0 |

## Chi-Square Calculations

In order to test the hypotheses discussed in Chapter 4, chi-square goodness-of-fit tests were performed on the survey data. Table C-4 shows the actual counts of participants' responses on the forced-choice questions, along with the outcomes of the goodness-of-fit test. The null hypothesis for this test claims that responses have a proportionally uniform distribution of [0.5, 0.5] across the choices, while the alternate hypothesis claims that responses are not consistent with that distribution. Therefore, the expected counts for a uniform distribution [4, 4] are included in the table apportioned equally between the two choice columns.

The goodness-of-fit test is performed on each row. The chi-square test statistic is calculated by the following formula:

$$\chi^2_{test} = \sum_{i=1}^{c} \frac{(O_i - E_i)^2}{E_i}$$

where $O_i$ is the observed value in column (i), $E_i$ is the expected value in column (i), and $c$ is the number of possible columns (two in this case). For example, the chi-square test statistic for Intel-0 is calculated by: (6 - 4)^2 / 4 + (2 - 4)^2 / 4 = 2. The critical chi-square values are retrieved from lookup tables and depend on the degrees of freedom for the problem and the desired level of alpha. In a goodness-of-fit test, the degrees of freedom are one number less than the number of possible columns (*df = c – 1)*. An alpha of 0.05 was used to reject the null hypothesis.

Table C-4. Chi-Square Goodness of Fit Tests, Forced Choice Questions

| General Question | Observed Count (n = 8) STPA | Traditional | Test Calculations (df = 1) $\chi^2$ crit ($\alpha$ = 0.05) | $\chi^2$ test statistic | p-value | |
|---|---|---|---|---|---|---|
| Intel-0 | 6 | 2 | 3.84 | 2.00 | 0.157 | >>>> Fail: $\chi^2$ (1, n = 8) = 2.00, p = 0.157 |
| Inform-0 | 7 | 1 | 3.84 | 4.50 | 0.034 | >> Reject: $\chi^2$ (1, n = 8) = 4.50, p < 0.05 |
| Implem-0 | 7 | 1 | 3.84 | 4.50 | 0.034 | >> Reject: $\chi^2$ (1, n = 8) = 4.50, p < 0.05 |
| *Expected Count* | *4* | *4* | | | | |

Table C-5. Chi-Square Goodness of Fit Tests, Detailed Questions

| Detailed Question | Observed Count (n = 8) STPA | Traditional | Both | Neither | Test Calculations (df = 3) $\chi^2$ crit ($\alpha$ = 0.05) | $\chi^2$ test statistic | p-value | |
|---|---|---|---|---|---|---|---|---|
| Intel-1 | 5 | 1 | 2 | 0 | 7.82 | 4.67 | 0.198 | >>>> Fail: $\chi^2$ (3, n = 8) = 4.67, p = 0.198 |
| Intel-2 | 4 | 2 | 2 | 0 | 7.82 | 2.67 | 0.446 | >>>> Fail: $\chi^2$ (3, n = 8) = 2.67, p = 0.446 |
| Intel-3 | 5 | 2 | 1 | 0 | 7.82 | 2.67 | 0.446 | >>>> Fail: $\chi^2$ (3, n = 8) = 2.67, p = 0.446 |
| Intel-4 | 7 | 0 | 1 | 0 | 7.82 | 9.33 | 0.025 | >> Reject: $\chi^2$ (3, n = 8) = 9.33, p < 0.05 |
| Intel-5 | 5 | 1 | 2 | 0 | 7.82 | 4.67 | 0.198 | >>>> Fail: $\chi^2$ (3, n = 8) = 4.67, p = 0.198 |
| Intel-6 | 6 | 0 | 2 | 0 | 7.82 | 8.00 | 0.046 | >> Reject: $\chi^2$ (3, n = 8) = 8.00, p < 0.05 |
| Intel-7 | 6 | 0 | 2 | 0 | 7.82 | 8.00 | 0.046 | >> Reject: $\chi^2$ (3, n = 8) = 8.00, p < 0.05 |
| Inform-1 | 8 | 0 | 0 | 0 | 7.82 | 13.33 | 0.004 | >> Reject: $\chi^2$ (3, n = 8) = 13.33, p < 0.01 |
| Inform-2 | 8 | 0 | 0 | 0 | 7.82 | 13.33 | 0.004 | >> Reject: $\chi^2$ (3, n = 8) = 13.33, p < 0.01 |
| Inform-3 | 5 | 2 | 1 | 0 | 7.82 | 2.67 | 0.446 | >>>> Fail: $\chi^2$ (3, n = 8) = 2.67, p = 0.446 |
| Inform-4 | 4 | 2 | 2 | 0 | 7.82 | 2.67 | 0.446 | >>>> Fail: $\chi^2$ (3, n = 8) = 2.67, p = 0.446 |
| Inform-5 | 8 | 0 | 0 | 0 | 7.82 | 13.33 | 0.004 | >> Reject: $\chi^2$ (3, n = 8) = 13.33, p < 0.01 |
| Inform-6 | 7 | 1 | 0 | 0 | 7.82 | 8.67 | 0.034 | >> Reject: $\chi^2$ (3, n = 8) = 8.67, p < 0.05 |
| Implem-1 | 4 | 2 | 2 | 0 | 7.82 | 2.67 | 0.446 | >>>> Fail: $\chi^2$ (3, n = 8) = 2.67, p = 0.446 |
| Implem-2 | 4 | 2 | 1 | 1 | 7.82 | 0.67 | 0.881 | >>>> Fail: $\chi^2$ (3, n = 8) = 0.67, p = 0.881 |
| Implem-3 | 4 | 4 | 0 | 0 | 7.82 | 2.67 | 0.446 | >>>> Fail: $\chi^2$ (3, n = 8) = 2.67, p = 0.446 |
| Implem-4 | 5 | 3 | 0 | 0 | 7.82 | 3.33 | 0.343 | >>>> Fail: $\chi^2$ (3, n = 8) = 3.33, p = 0.343 |
| Implem-5 | 7 | 0 | 1 | 0 | 7.82 | 9.33 | 0.025 | >> Reject: $\chi^2$ (3, n = 8) = 9.33, p < 0.05 |
| Implem-6 | 5 | 2 | 0 | 1 | 7.82 | 2.67 | 0.446 | >>>> Fail: $\chi^2$ (3, n = 8) = 2.67, p = 0.446 |
| Implem-7 | 5 | 0 | 3 | 0 | 7.82 | 9.33 | 0.025 | >> Reject: $\chi^2$ (3, n = 8) = 9.33, p < 0.05 |
| *Expected Count* | *3* | *3* | *1* | *1* | | | | |

Additionally, the p-value for each statistic can be calculated using reverse lookup tables, again depending on the degrees of freedom. If the test statistic is greater than the critical value for a given alpha, the p-value will be less than that alpha value. Calculating the specific p-values is useful, as it gives the investigator the specific probability that the observed results were random (i.e., the probability of incorrectly rejecting the null when it is actually true).

Table C-5 shows the counts and goodness-of-fit outcomes for the responses to the twenty detailed questions. The degrees of freedom for each test are updated to reflect the existence of four choices. The goodness-of-fit hypotheses for the detailed questions are slightly different than for the forced-choice questions. The null hypothesis claims that responses have a proportional distribution of [0.375, 0.375, 0.125, 0.125] across the choices, while the alternate hypothesis claims that responses are not consistent with that distribution. Therefore, the counts for the expected distribution [3, 3, 1, 1] are included in the table apportioned equally between the four choice columns.

## Short Answer Outputs

The following is a collection of the short answers received from the eight survey participants, edited only for clarity and to remove jargon and sensitive data. Not all the short-answer fields were mandatory, and some participants made multiple points, so some fields contain more statements than others.

### *What do you like the best about each method?*

Hazard analysis is more straightforward with STPA.

STPA: clear distinction between an accident and a hazard.

The STPA method actually analyzes the risk inherent in a whole system. It appears that STPA would be much more effective at determining the true risks involved in testing.

I like the trickle down traceability of the STPA approach.

STPA clearly structures the analysis, from system description and it's boundary with the environment, to recovery procedures.

The STPA format was much easier to follow and provided much more usable information

The STPA is absolutely a more accurate, explicit, clear picture of the "situation" overall, which results in a better product for risk mitigation.

Traditional is comfortable, because it is what I know best.

Traditional: simple single hazards called out that will be easy to keep in mind during flight.

More familiarity with traditional approach.

The traditional safety plan gives decision makers simple data points on which to base decisions (risk level) and simple things to implement during execution (brief GMCs and THAs).

The legacy report unfortunately caters to AFMC's favorite things: speed, convenience, and the reuse of old test plans to write new ones.

Both present the safety concerns clearly.

### *What do you like the least about each?*

The Achilles heel of STPA method is also what makes it strong: it requires thought and time. I'm afraid this will be why lazy decision-makers will avoid it.

STPA is hard to say, haven't implemented it yet.

The STPA included too much info; someone unfamiliar with a safety plan would have difficulty navigating to find the desired info.

STPA - While a well thought out and written plan would likely be far more effective at identifying and reducing risk, the actual implementation would probably negate all benefits because of the time consuming and very intricate nature of the analysis. Unless a "safety engineering" function were stood up and fully staffed the safety personnel would likely not

have time or spare mental capacity to properly execute the analysis. Further, the audience (testers and decision makers) would have to devote far more time to fully grasp the details of the STPA method.

STPA - I got lost in the parenthetical cross-references during the system description.

STPA: the complicated control analysis.

STPA requires greater management guidance up front... The product content is sensitive to system definition, for example.

The traditional safety method is unsatisfactory due to its lazy approach of simply copy/pasting the last one, then continuing the linear cause/effect style of analysis originating solely from whatever may occur to a group of planners. It's riddled with dangerous holes in the analyses.

The traditional method devolves to selecting a package of THAs and GMCs, without providing the underlying analysis.

Traditional: GMC/THAs. Mission briefings are way too long and repetitive. After more than one mission using the same safety package, crew members become bored and complacent.

Traditional - The actual effectiveness of the safety plan is probably marginal as the authors of these plans typically heavily leverage old plans as examples and don't necessarily have a full understanding of the system. Additionally, reading GMCs and THAs in the pre-mission briefing is often of zero use because people tune it out...especially if they have heard it multiple times in a single week.

Traditional: unclear about what belongs in the safety plan from the technical plan.

The traditional method can provide negative training value, by creating bad safety planning habits.

Both plans can easily over constrain the test team by placing undue measures of mitigation before the problem is fully understood.

### *How much time would you recommend to someone for learning the basics of each?*

Teaching STPA is similar to teaching someone how to use a TI-89 calculator versus a four-function solar calculator. The former will take significantly longer to teach than the latter, but the investment is worth so much more.

3-4 hours of classroom, plus a practical app exercise.

I would say it takes me a read through of a plan and a discussion to learn the basics. That would take about an hour per plan for the basics.

Traditional = 3 months, STPA = 1 month.

Traditional - six months to a year of apprenticeship under a qualified safety officer with at least 3-4 safety packages written during that time. STPA - 1-2 years of apprenticeship under

a qualified STPA safety officer. Additionally, STPA safety officer candidates would probably be best pulled from a systems-engineering background instead of operations.

### *Which method would you prefer to use for your next test project, and why?*

STPA [no further comment].

STPA, the older method is not taught well, nor understood what actually belongs in each section.

STPA. Important to get safety planning right, despite the larger up-front investment.

I would like to use the STPA for the clear distinction between accidents and hazards. I also like the ease of going through a list and selecting the applicable accidents and hazards, but I'm concerned that I might miss something without brainstorming these first.

Assuming I had the time to implement it, I would absolutely like to use the STPA method. I'd be interested to see which "classic" items are diminished and what new considerations arise.

Traditional, because I know the process and could likely execute it in the time available during the project. However, I would love to have access to the information contained in the STPA method.

### *Do you have any suggestions for the formatting and information ordering in the STPA planning document?*

Put only what is actually required for safety. It must be written hand-in-hand with the technical plan, or it is very easy to repeat oneself. Going to the STPA format suggests modifications to the technical-plan format that might be required as well.

Keep the mitigation to a single page and don't expound on things everyone in the test business knows.

No. Its layout is logical and satisfactory.

### *Additional Comments*

Good luck, can't wait to see you change the test enterprise!

Gotta get the new Test Center commander and Test Wing commander on board.

The STPA format will require more thought to both write and comprehend, but I would argue that's the level of mental effort that safety planning requires and is currently being deprived.

# References


[1] T. Wolfe, *The Right Stuff*. Farrar, Straus and Giroux, 1979.

[2] N. N. Taleb, *The Black Swan: The Impact of the Highly Improbable Fragility*. Random House LLC, 2010.

[3] T. Gilovich, D. Griffin, and D. Kahneman, Eds., *Heuristics and Biases: The Psychology of Intuitive Judgment*, 1st ed. Cambridge University Press, 2002.

[4] S. W. A. Dekker, *The Field Guide to Understanding Human Error*, 1st ed. Ashgate Publishing, Ltd., 2006.

[5] N. G. Leveson, *Engineering a Safer World: Systems Thinking Applied to Safety*. The MIT Press, 2012.

[6] E. L. Trist and K. W. Bamforth, "Some social and psychological consequences of the Longwall method," *Human Relations*, vol. 4, no. 3, pp. 3–38, 1951.

[7] J. Rasmussen, "Risk management in a dynamic society: A modelling problem," *Safety Science*, vol. 27, no. 2, pp. 183–213, 1997.

[8] J. Rasmussen, A. M. Pejtersen, and L. P. Goodstein, *Cognitive Systems Engineering*, 1st ed. Wiley-Interscience, 1994.

[9] N. G. Leveson, "Intent specifications: An approach to building human-centered specifications," *IEEE Transactions on Software Engineering*, vol. 26, no. 1, pp. 15–35, 2000.

[10] J. Rasmussen and L. P. Goodstein, "Decision support in supervisory control," Risø National Laboratoty, Denmark, RISO-M-2525, 1985.

[11] "MIL-STD-882E: System Safety." Department of Defense, 2012.

[12] E. Wenger, *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, 1999.

[13] "Title 41, Part 101-26: Procurement Sources and Program." Code of Federal Regulations, 2015.

[14] J. Noyes, "The QWERTY keyboard: A review," *International Journal of Man-Machine Studies*, vol. 18, no. 3, pp. 265–281, 1983.

[15] *Autonomy Science and Technology Strategy*. Air Force Research Laboratory, 2013.

[16] P. W. Singer, *Wired for War: The Robotics Revolution and Conflict in the 21st Century*, Reprint. Penguin Books, 2009.

[17] "Circular 328 AN/190: Unmanned Aircraft Systems (UAS)." International Civil Aviation Organization, 2011.

[18] *Unmanned Aircraft Systems Roadmap 2010-2035*. U.S. Army, UAS Center of Excellence, 2010.

[19] C. Whitlock, "When drones fall from the sky," *Washington Post*, 2014. [Online]. Available: http://www.washingtonpost.com/sf/investigative/2014/06/20/when-drones-fall-from-the-sky.

[20] *Integration of Civil Unmanned Aircraft Systems in the National Airspace System Roadmap*. Federal Aviation Administration, 2013.

[21] M. V. Stringfellow, “Accident analysis and hazard analysis for human and organizational factors,” Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2011.
[22] G. Dobie, “Global aviation safety study,” Allianz Global Corporate and Specialty, 2014.
[23] *Defense Acquisition Guidebook*. Department of Defense, 2013.
[24] “Air Force Test Center Instruction 91-203: AFTC Test Safety Review Policy.” 2014.
[25] “Title 14, Part 91: General Operating and Flight Rules.” Code of Federal Regulations, 2014.
[26] “Edwards Air Force Base Instruction 13-100: Flying and Airfield Operations.” 2013.
[27] “Range Safety Group Standard 321-10: Common Risk Criteria Standards for National Test Ranges.” Range Commander’s Council, Department of Defense, 2010.
[28] “MIL-HDBK-516C: Airworthiness Certification Criteria.” Department of Defense, 2014.
[29] “Air Force Instruction 91-202: Mishap Prevention Program.” 2013.
[30] “Air Force Instruction 90-802: Risk Management.” 2013.
[31] “Air Force Instruction 11-200: Aircrew Training, Standardization/Evaluation, and General Operations Structure.” 2012.
[32] “Air Force Instruction 13-204v3: Airfield Operations Procedures and Programs.” 2012.
[33] “Air Force Instruction 13-201: Airspace Management.” 2012.
[34] “Air Force Instruction 13-212: Range Planning and Operations.” 2010.
[35] “Air Force Instruction 91-204: Safety Investigations and Reports.” 2014.
[36] “Air Force Instruction 11-290: Cockpit/Crew Resource Management Program.” 2012.
[37] P. W. Merlin, G. A. Bendrick, and D. A. Holland, *Breaking the Mishap Chain: Human Factors Lessons Learned From Aerospace Accidents and Incidents in Research, Flight Test, and Development*. NASA Aeronautics Book Series, 2012.
[38] N. G. Leveson, L. D. Pinnel, S. D. Sandys, S. Koga, and J. D. Reese, “Analyzing software specifications for mode confusion potential,” in *Proceedings of a Workshop on Human Error and System Development*, Glasgow, Scotland, 1997, pp. 132–146.
[39] J. J. Clark and R. K. Goulder, “Human systems integration (HSI): Ensuring design and development meet human performance capability early in acquisition process,” *Program Manager*, vol. 34, no. 9, pp. 88–91, Aug-2002.
[40] “NASA human systems integration division overview,” 2014. [Online]. Available: http://human-factors.arc.nasa.gov/awards_pubs/factsheets.php.
[41] *Human Systems Integration Handbook*. U.S. Air Force, 711 HPW/HPO, 2009.
[42] *Human Systems Integration Requirements Pocket Guide*. U.S. Air Force, SAF/AQ-AFHSIO, 2009.
[43] E. H. Schein, *Organizational Culture and Leadership*, 4th ed. Jossey-Bass, 2010.

[44] H. Beyer and K. Holtzblatt, *Contextual Design: Defining Customer-Centered Systems*, 1st ed. Morgan Kaufmann, 1997.
[45] P. Checkland, *Systems Thinking, Systems Practice: Includes a 30-year Retrospective*. Wiley, 1999.
[46] S. J. Kapurch, *NASA Systems Engineering Handbook*. DIANE Publishing, 2007.
[47] A. M. Madni, "Integrating humans with and within complex systems: Challenges and opportunities," *CrossTalk*, vol. 24, no. 3, pp. 4–8, Jun-2011.
[48] C. A. Ericson II, "A short history of system safety," *Journal of System Safety*, 2006. [Online]. Available: http://www.system-safety.org/ejss/past/novdec2006ejss/clifs.php.
[49] J. M. Flach, J. S. Carroll, M. J. Dainoff, and W. I. Hamilton, "Striving for safety: Communicating and deciding in sociotechnical systems," *Ergonomics*, pp. 1–20, Mar. 2015.
[50] J. C. Smuts, *Holism and evolution*. London: McMillan and Co. Limited, 1926.
[51] L. von Bertalanffy, *General System Theory: Foundations, Development, Applications*. Braziller, New York, 1968.
[52] N. Wiener, *Cybernetics or Control and Communication in the Animal and the Machine*. Paris: Hermann, 1948.
[53] N. G. Leveson, *Safeware: System Safety and Computers*, 1st ed. Addison-Wesley Professional, 1995.
[54] G. E. P. Box and N. R. Draper, *Empirical Model-Building and Response Surfaces*. Oxford, England: John Wiley & Sons, 1987.
[55] J. D. Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin/McGraw-Hill Boston, 2000.
[56] S. L. Star and J. R. Griesemer, "Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39," *Social studies of science*, vol. 19, no. 3, pp. 387–420, 1989.
[57] B. A. Bechky, "Sharing Meaning Across Occupational Communities: The Transformation of Understanding on a Production Floor," *Organization Science*, vol. 14, no. 3, pp. 312–330, Jun. 2003.
[58] R. W. Proctor and T. V. Zandt, *Human Factors in Simple and Complex Systems*, 2nd ed. CRC Press, 2008.
[59] C. D. Wickens, J. G. Hollands, R. Parasuraman, and S. Banbury, *Engineering Psychology and Human Performance*, 4th ed. Pearson, 2012.
[60] T. B. Sheridan, *Humans and Automation: System Design and Research Issues*, 1st ed. Wiley-Interscience, 2002.
[61] F. W. Taylor, *The Principles of Scientific Management*. Harper, 1914.
[62] J. M. Flach, F. Tanabe, K. Monta, K. J. Vicente, and J. Rasmussen, "An ecological approach to interface design," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1998, vol. 42, pp. 295–299.
[63] C. C. Gordon, T. Churchill, C. E. Clauser, B. Bradtmiller, and J. T. McConville, "Anthropometric Survey of U.S. Army Personnel: Methods and Summary Statistics 1988," Sep. 1989.
[64] F. B. Gilbreth and E. G. Carey, *Cheaper by the Dozen*. Martino Fine Books, 1948.

[65] L. R. Young, “Spatial orientation,” in *Principles and Practice of Aviation Psychology*, P. S. Tsang and M. A. Vidulich, Eds. Lawrence Erlbaum Associates Publishers, 2003, pp. 69–113.
[66] J. C. Buckey, *Space Physiology*. Oxford University Press, USA, 2006.
[67] K. E. Klein and H. M. Wegmann, “Circadian rhythms in air operations,” in *AGARD Sleep, Wakefulness and Circadian Rhythm*, 1979.
[68] C. Borst, J. M. Flach, and J. Ellerboek, “Beyond ecological interface design: Lessons from concerns and misconceptions,” *IEEE Transactions on Human-Machine Systems*, vol. 45, no. 2, pp. 164–175, 2014.
[69] J. M. Flach, “Situation awareness: Context matters!: A commentary on Endsley,” *Journal of Cognitive Engineering and Decision Making*, vol. 9, no. 1, pp. 59–72, 2015.
[70] P. M. Fitts, “Human engineering for an effective air-navigation and traffic-control system,” National Research Council, Washington, D.C., 1951.
[71] J. C. F. de Winter and D. Dodou, “Why the Fitts list has persisted throughout the history of function allocation,” *Cognition, Technology & Work*, vol. 16, no. 1, pp. 1–11, Aug. 2011.
[72] D. T. McRuer, “Human pilot dynamics in compensatory systems,” US Flight Dynamics Laboratory, 1965.
[73] L. R. Young, “Human control capabilities,” in *Bioastronautics Data Book, 2nd Edition*, J. F. Parker Jr. and V. R. West, Eds. NASA, 1973, pp. 751–759.
[74] T. B. Sheridan and W. R. Ferrell, *Man-Machine Systems: Information, Control, and Decision Models of Human Performance*. The MIT Press, 1974.
[75] R. J. Jagacinski and J. M. Flach, *Control Theory for Humans: Quantitative Approaches To Modeling Performance*, 1st ed. CRC Press, 2002.
[76] T. B. Sheridan and W. L. Verplank, “Human and Computer Control of Undersea Teleoperators,” Office of Naval Research, N00014-77-C-0256, Jul. 1978.
[77] T. B. Sheridan, *Telerobotics, Automation, and Human Supervisory Control*. The MIT Press, 2003.
[78] T. B. Sheridan, “Supervisory control,” in *Handbook of Human Factors and Ergonomics*, G. Salvendy, Ed. John Wiley & Sons, Inc., 2006, pp. 1025–1052.
[79] M. R. Endsley and D. B. Kaber, “Level of automation effects on performance, situation awareness and workload in a dynamic control task,” *Ergonomics*, vol. 42, no. 3, pp. 462–492, 1999.
[80] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “A model for types and levels of human interaction with automation,” *IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans*, vol. 30, no. 3, pp. 286–297, 2000.
[81] J. M. Flach, P. F. Jacques, D. L. Patrick, M. Amelink, M. M. Van Paassen, and M. Mulder, “A search for meaning: A case study of the approach-to-landing,” in *Handbook of Cognitive Task Design*, CRC Press, 2003, pp. 171–191.
[82] M. H. J. Amelink, M. Mulder, R. M. M. Van Paassen, and J. Flach, “Theoretical foundations for a total energy-based perspective flight-path display,” *The International Journal of Aviation Psychology*, vol. 15, no. 3, pp. 205–231, 2005.

[83] J. J. Abbott, P. Marayong, and A. M. Okamura, “Haptic virtual fixtures for robot-assisted manipulation,” in *Robotics Research: Results of the 12th International Symposium ISRR*, vol. 28, Springer, 2007, pp. 49–64.
[84] S. Y. Nof, *Springer Handbook of Automation*. Springer Science & Business Media, 2009.
[85] E. L. Wiener and D. C. Nagel, Eds., *Human Factors in Aviation*, 1st ed. Academic Press, 1989.
[86] M. Rogovin, “Three Mile Island: A report to the commissioners and to the public,” Nuclear Regulatory Commission, Washington, DC (USA), NUREG/CR-1250(Vol.1), Jan. 1979.
[87] N. B. Sarter and D. D. Woods, “How in the world did we ever get into that mode? Mode error and awareness in supervisory control,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 5–19, 1995.
[88] N. B. Sarter, D. D. Woods, and C. E. Billings, “Automation surprises,” in *Handbook of Human Factors and Ergonomics*, vol. 2, John Wiley and Sons, 1997, pp. 1926–1943.
[89] N. B. Sarter and D. D. Woods, “Team play with a powerful and independent agent: A full-mission simulation study,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 42, no. 3, pp. 390–402, Sep. 2000.
[90] S. W. A. Dekker and D. D. Woods, “MABA-MABA or abracadabra? Progress on human-automation coordination,” *Cognition, Technology & Work*, vol. 4, no. 4, pp. 240–244, Nov. 2002.
[91] N. G. Leveson and E. Palmer, “Designing automation to reduce operator errors,” in *IEEE International Conference on Systems, Man, and Cybernetics*, 1997, vol. 2, pp. 1144–1150.
[92] J. B. Watson, “Psychology as the behaviorist views it,” *Psychological Review*, vol. 20, no. 2, pp. 158–177, 1913.
[93] P. A. Ertmer and T. J. Newby, “Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective,” *Performance Improvement Quarterly*, vol. 26, no. 2, pp. 43–71, 2013.
[94] P. M. Fitts, “The information capacity of the human motor system in controlling the amplitude of movement,” *Journal of Experimental Psychology*, vol. 47, no. 6, pp. 381–391, 1954.
[95] J. R. Simon, “Reactions toward the source of stimulation,” *Journal of Experimental Psychology*, vol. 81, no. 1, pp. 174–176, 1969.
[96] U. Neisser, *Cognitive Psychology*. Englewood Cliffs, N.J: Prentice-Hall, Inc., 1967.
[97] G. A. Miller, “The magical number seven, plus or minus two: Some limits on our capacity for processing information,” *Psychological Review*, vol. 63, no. 2, pp. 81–97, 1956.
[98] R. C. Atkinson and R. M. Shiffrin, “Human memory: A proposed system and its control processes,” in *Psychology of Learning and Motivation*, vol. 2, K. W. Spence and J. Taylor, Eds. Academic Press, 1968, pp. 89–195.
[99] A. D. Baddeley and G. Hitch, “Working memory,” in *Psychology of Learning and Motivation*, vol. 8, G. H. Bower, Ed. Academic Press, 1974, pp. 47–89.

[100] A. T. Welford, “The ‘Psychological Refractory Period’ and the timing of high-speed performance: A review and a theory,” *British Journal of Psychology*, vol. 43, no. 1, pp. 2–19, Feb. 1952.
[101] C. D. Wickens, “Multiple resources and mental workload,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 50, no. 3, pp. 449–455, 2008.
[102] R. M. Yerkes and J. D. Dodson, “The relation of strength of stimulus to rapidity of habit-formation,” *J. Comp. Neurol. Psychol.*, vol. 18, no. 5, pp. 459–482, Nov. 1908.
[103] P. L. Broadhurst, “Emotionality and the Yerkes-Dodson Law,” *Journal of Experimental Psychology*, vol. 54, no. 5, pp. 345–352, 1957.
[104] J. Y. C. Chen, M. J. Barnes, and M. Harper-Sciarini, “Supervisory control of multiple robots: Human-performance issues and user-interface design,” *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 41, no. 4, pp. 435–454, Jul. 2011.
[105] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs,” *Journal of Cognitive Engineering and Decision Making*, vol. 2, no. 2, pp. 140–160, Jul. 2008.
[106] G. E. Cooper and R. P. Harper Jr., “The use of pilot rating in the evaluation of aircraft handling qualities,” NASA, TN D-5153, 1969.
[107] M. L. Cummings, K. Myers, and S. D. Scott, “Modified Cooper Harper evaluation tool for unmanned vehicle displays,” in *Conference on Unmanned Vehicle Systems*, Canada, 2006.
[108] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research,” in *Advances in Psychology*, vol. 52, P. A. H. and N. Meshkati, Ed. North-Holland, 1988, pp. 139–183.
[109] G. B. Reid and T. E. Nygren, “The Subjective Workload Assessment Technique: A scaling procedure for measuring mental workload,” in *Advances in Psychology*, vol. 52, P. A. H. and N. Meshkati, Ed. North-Holland, 1988, pp. 185–218.
[110] M. R. Endsley, “Toward a theory of situation awareness in dynamic systems,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 37, no. 1, pp. 32–64, Mar. 1995.
[111] M. R. Endsley, “Situation awareness in aviation systems,” in *Handbook of Aviation Human Factors*, Lawrence Erlbaum Associates Mahwah, NJ, 1999, pp. 257–276.
[112] F. T. Durso, A. R. Dattel, S. Banbury, and S. Tremblay, “SPAM: The real-time assessment of SA,” in *A Cognitive Approach to Situation Awareness: Theory, Measures and Application*, Ashgate Publishing, Ltd., 2004, pp. 137–154.
[113] M. R. Endsley, “Situation awareness global assessment technique (SAGAT),” in *IEEE National Aerospace and Electronics Conference*, 1988, pp. 789–795 vol.3.
[114] J. D. Lee and K. A. See, “Trust in automation: Designing for appropriate reliance,” *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 46, no. 1, pp. 50–80, Jan. 2004.

[115] C. A. Miller, "Trust in adaptive automation: The role of etiquette in tuning trust via analogic and affective methods," in *Proceedings of the 1st International Conference on Augmented Cognition*, Las Vegas, NV, 2005.
[116] M. T. Dzindolet, S. A. Peterson, R. A. Pomranky, L. G. Pierce, and H. P. Beck, "The role of trust in automation reliance," *International Journal of Human-Computer Studies*, vol. 58, no. 6, pp. 697–718, Jun. 2003.
[117] R. Parasuraman and V. Riley, "Humans and automation: Use, misuse, disuse, abuse," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 39, no. 2, pp. 230–253, Jun. 1997.
[118] J. Riegelsberger, M. A. Sasse, and J. D. McCarthy, "The mechanics of trust: A framework for research and design," *International Journal of Human-Computer Studies*, vol. 62, no. 3, pp. 381–422, 2005.
[119] M. Steinberg, "Moving from Supervisory Control of Autonomous Systems to Human-Machine Teaming," presented at the 4th Annual Human-Agent-Robot Teamwork Workshop, 2012.
[120] A. S. Clare, "Modeling real-time human-automation collaborative scheduling of unmanned vehicles," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2013.
[121] "MIL-HDBK-1908B: Definitions of Human Factors Terms." Department of Defense, 1999.
[122] "MIL-STD-1472G: Human Engineering." Department of Defense, 2012.
[123] "MIL-STD-46855A: Human Engineering Requirements for Military Systems, Equipment, and Facilities." Department of Defense, 2011.
[124] "MIL-HDBK-87213A: Electronically/Optically Generated Airborne Displays." Department of Defense, 2005.
[125] "MIL-STD-1787C: Aircraft Display Symbology." Department of Defense, 2001.
[126] "MIL-STD-411F: Aircrew Station Alerting Systems." Department of Defense, 1997.
[127] "MIL-STD-1797A: Flying Qualities of Piloted Aircraft." Department of Defense, 1995.
[128] "MIL-STD-1474D: Noise Limits." Department of Defense, 1997.
[129] W. R. Ashby, *An Introduction to Cybernetics*. London: Methuen, 1956.
[130] J. M. Histon, "Mitigating complexity in air traffic control: the role of structure-based abstractions," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2008.
[131] S. W. A. Dekker and E. Hollnagel, "Human factors and folk models," *Cogn Tech Work*, vol. 6, no. 2, pp. 79–86, Oct. 2003.
[132] J. M. Flach, "The concept of the situation in psychology," in *A Cognitive Approach To Situation Awareness: Theory And Application*, S. P. Banbury and S. Tremblay, Eds. Ashgate Publishing, Ltd., 2004.
[133] E. Brunswik, *Perception and the Representative Design of Psychological Experiments*. University of California Press, 1956.
[134] B. H. Kantowitz and R. D. Sorkin, *Human Factors: Understanding People-System Relationships*, 1st ed. New York: Wiley, 1983.

[135] S. W. A. Dekker, *Ten Questions about Human Error: A New View of Human Factors and System Safety*. CRC Press, 2004.
[136] N. G. Leveson, J. Cutcher-Gershenfeld, B. Barrett, A. Brown, J. Carroll, N. Dulac, L. Fraile, and K. Marais, "Effectively addressing NASA's organizational and safety culture: Insights from systems safety and engineering systems," in *Engineering Systems Symposium*, MIT, 2004.
[137] "National Policy Order 8130.2H: Airworthiness Certification of Products and Articles." Federal Aviation Administration, Feb-2015.
[138] "Department of Defense Directive 5030.61: Airworthiness Policy." 2013.
[139] "MIL-P-1629: Procedures for Performing a Failure Mode Effect and Critical Analysis." Department of Defense, 1949.
[140] A. F. Hixenbaugh, "Fault Tree for Safety," Boeing Co. Support Systems Engineering, D6-53604, Nov. 1968.
[141] J. Reason, "The contribution of latent human failures to the breakdown of complex systems," *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, vol. 327, no. 1241, pp. 475–484, Apr. 1990.
[142] J. E. Gordon, "Epidemiology in modern perspective," *Proceedings of the Royal Society of Medicine*, vol. 47, no. 7, pp. 564–570, Jul. 1954.
[143] A. Rae, J. McDermid, and R. Alexander, "The science and superstition of quantitative risk assessment," *Journal of Systems Safety*, vol. 48, no. 4, p. 28, 2012.
[144] M. Hazewinkel, "Law of large numbers," in *Encyclopedia of Mathematics*, Berlin: Springer, 2001.
[145] D. Kahneman, *Thinking, Fast and Slow*, 1st ed. Farrar, Straus and Giroux, 2011.
[146] S. Pruchnicki, "Top ten ways to destroy your just culture," in *Proceedings of the 18th International Symposium on Aviation Psychology*, 2015.
[147] H. P. Ginsburg and S. Opper, *Piaget's Theory of Intellectual Development*. Prentice-Hall, Inc, 1988.
[148] W. F. Brewer and G. V. Nakamura, "The nature and functions of schemas," Champaign, Ill.: University of Illinois at Urbana-Champaign, Center for the Study of Reading., 1984.
[149] W. G. Perry, "Cognitive and ethical growth: The making of meaning," *Arthur W. Chickering and Associates, The Modern American College*, vol. 4, pp. 48–116, 1981.
[150] G. I. Rochlin, "Safe operation as a social construct," *Ergonomics*, vol. 42, no. 11, pp. 1549–1560, Nov. 1999.
[151] C. S. Peirce, *Philosophical Writings of Peirce*. Courier Corporation, 2012.
[152] A. Kirlik, "Requirements for psychological models to support design: Towards ecological task analysis," in *Global Perspectives on the Ecology of Human-Machine Systems*, vol. 1, J. M. Flach, P. A. Hancock, J. E. Caird, and K. J. Vicente, Eds. Lawrence Erlbaum Associates, Inc., 1995, pp. 68–120.
[153] J. M. Flach, "What Matters?," Dayton, OH, Unpublished manuscript-2015.
[154] J. J. Gibson, *The Ecological Approach To Visual Perception*, New Ed. Psychology Press, 1986.
[155] L. Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*. Cambridge University Press, 2007.

[156] E. Hutchins, *Cognition in the Wild*, New edition. A Bradford Book, 1996.
[157] D. D. Woods and E. Hollnagel, *Joint Cognitive Systems: Patterns in Cognitive Systems Engineering*. CRC Press, 2006.
[158] E. Hutchins, "How a cockpit remembers its speeds," *Cognitive Science*, vol. 19, no. 3, pp. 265–288, 1995.
[159] E. Hutchins and T. Klausen, "Distributed cognition in an airline cockpit," in *Cognition and Communication at Work*, Cambridge University Press, New York, NY, 1998, pp. 15–34.
[160] N. A. Stanton, P. M. Salmon, G. H. Walker, and D. P. Jenkins, "Is situation awareness all in the mind?," *Theoretical Issues in Ergonomics Science*, vol. 11, no. 1–2, pp. 29–40, 2010.
[161] C. E. Lindblom, "Still muddling, not yet through," *Public Administration Review*, vol. 39, no. 6, pp. 517–526, 1979.
[162] K. E. Weick, *Sensemaking in Organizations*, vol. 3. Sage, 1995.
[163] R. Lipshitz, G. Klein, J. Orasanu, and E. Salas, "Taking stock of naturalistic decision making," *Journal of Behavioral Decision Making*, vol. 14, no. 5, pp. 331–352, Dec. 2001.
[164] G. Klein, *Streetlights and Shadows: Searching for the Keys to Adaptive Decision Making*. A Bradford Book, 2011.
[165] J. Rasmussen, "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-13, no. 3, pp. 257–266, 1983.
[166] J. Rasmussen, *Information Processing and Human-Machine Interaction. An Approach to Cognitive Engineering*, vol. 12. Elsevier Science Ltd, 1986.
[167] G. Lintern, "Cognitive work analysis," *Cognitive Systems Design*, 2013. [Online]. Available: http://cognitivesystemsdesign.net/Tutorials/.
[168] G. Lintern, "Work domain analysis," *Cognitive Systems Design*, 2013. [Online]. Available: http://cognitivesystemsdesign.net/Tutorials.
[169] N. A. Stanton, "Hierarchical task analysis: Developments, applications, and extensions," *Applied Ergonomics*, vol. 37, no. 1, pp. 55–79, Jan. 2006.
[170] J. R. Hajdukiewicz, C. M. Burns, K. J. Vicente, and R. G. Eggleston, "Work domain analysis for intentional systems," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 1999, vol. 43, pp. 333–337.
[171] K. J. Vicente, *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*, 1st ed. CRC Press, 1999.
[172] C. A. Miller and K. J. Vicente, "Comparison of display requirements generated via hierarchical task and abstraction-decomposition space analysis techniques," *International Journal of Cognitive Ergonomics*, vol. 5, no. 3, pp. 335–355, 2001.
[173] N. Naikar, *Work Domain Analysis: Concepts, Guidelines, and Cases*. CRC Press, 2013.
[174] G. A. Klein, D. D. Woods, J. M. Bradshaw, R. R. Hoffman, and P. J. Feltovich, "Ten challenges for making automation a 'team player' in joint human-agent activity," *IEEE Intelligent Systems*, vol. 19, no. 6, pp. 91–95, 2004.
[175] P. W. Merlin, *Crash Course: Lessons Learned from Accidents Involving Remotely Piloted and Autonomous Aircraft*. NASA Aeronautics Book Series, 2013.

[176] J. Reason, "Human error: models and management," *BMJ*, vol. 320, no. 7237, pp. 768–770, 2000.
[177] N. Stanton and C. Baber, "A systems approach to human error identification," *Safety Science*, vol. 22, no. 1, pp. 215–228, 1996.
[178] N. A. Stanton, P. Salmon, D. Harris, A. Marshall, J. Demagalski, M. S. Young, T. Waldmann, and S. W. A. Dekker, "Predicting pilot error: Testing a new methodology and a multi-methods and analysts approach," *Applied ergonomics*, vol. 40, no. 3, pp. 464–471, 2009.
[179] N. A. Stanton and P. M. Salmon, "Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems," *Safety Science*, vol. 47, no. 2, pp. 227–237, 2009.
[180] B. Antoine, "Systems theoretic hazard analysis (STPA) applied to the risk review of complex systems: An example from the medical device industry," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2013.
[181] N. C. Dunn, "Satellite system safety analysis using STPA," M.S. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2013.
[182] J. Hickey, "A system theoretic safety analysis of US Coast Guard aviation mishap involving CG-6505," M.S. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2012.
[183] P. S. Nelson, "A STAMP analysis of the LEX COMAIR 5191 accident," M.S. Thesis, Lund University, Sweden, 2008.
[184] N. G. Leveson, C. Wilkinson, C. Fleming, J. Thomas, and I. Tracy, "A comparison of STPA and the ARP 4761 safety assessment process," Massachusetts Institute of Technology, NNL10AA13C-4.7.1, May 2014.
[185] C. H. Fleming, "Safety-driven early concept analysis and development," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2015.
[186] P. Y. Lipscy, K. E. Kushida, and T. Incerti, "The Fukushima disaster and Japan's nuclear plant vulnerability in comparative perspective," *Environmental Science and Technology*, vol. 47, no. 12, pp. 6082–6088, 2013.
[187] N. Polmar, *The Death of the USS Thresher: The Story Behind History's Deadliest Submarine Disaster*. Globe Pequot, 2004.
[188] N. Dulac, "A framework for dynamic safety and risk management modeling in complex engineering systems," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2007.
[189] J. P. Thomas IV, "Extending and automating a systems-theoretic hazard analysis for requirements generation and analysis," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2013.
[190] C. L. Thornberry, "Extending the human-controller methodology in systems-theoretic process analysis (STPA)," M.S. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2014.
[191] "Department of Defense Directive 5000.01: The Defense Acquisition System." 2007.
[192] "Department of Defense Directive 5000.02: Operation of the Defense Acquisition System." 2008.

[193] "Department of Defense Directive 5134.01: Under Secretary of Defense for Acquisition, Technology, and Logistics." 2008.
[194] "Department of Defense Directive 5141.02: Director of Operational Test and Evaluation." 2009.
[195] "Department of Defense Instruction 5134.16: Deputy Assistant Secretary of Defense for Systems Engineering." 2011.
[196] "Department of Defense Instruction 5134.17: Deputy Assistant Secretary of Defense for Developmental Test and Evaluation." 2011.
[197] "Department of Defense Directive 5105.71: Test Resource Management Center." 2004.
[198] "Department of Defense Directive 3200.11: Major Range and Test Facility Base." 2007.
[199] "Department of Defense Instruction 3200.18: Management and Operation of the Major Range and Test Facility Base." 2010.
[200] "Chairman of the Joint Chiefs of Staff Instruction 3170.01: Joint Capabilities Integration and Development System." Department of Defense, 2012.
[201] G. L. Feithans, A. J. Rowe, J. E. Davis, M. Holland, and L. Berger, "Vigilant Spirit control station (VSCS): The face of COUNTER," in *Proceedings of AIAA Guidance, Navigation and Control Conference*, 2008.
[202] M. Draper, G. Calhoun, H. Ruff, B. Mullins, A. Lefebvre, A. Ayala, and N. Wright, "Transition display aid for changing camera views in UAV operations," in *Proceedings of the First Conference on Humans Operating Unmanned Systems*, 2008.
[203] "Title 14, Parts 400–460: Commercial Space Transportation." Code of Federal Regulations, 2013.
[204] A. C. Horrell, "Extending safety assessment methods for remotely piloted aircraft operations in the national airspace system," M.S. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2010.
[205] D. Y. Cowsert, "Safety at center of growing RPA requirement," *Air Force News*, Sep-2012. [Online]. Available: http://www.af.mil/News/ArticleDisplay/tabid/223/Article/110490/safety-at-center-of-growing-rpa-requirement.aspx.
[206] *Joint UAS Concept of Operations, 2nd Edition*. Department of Defense, 2008.
[207] T. M. Cullen, "The MQ-9 Reaper remotely piloted aircraft: Humans and machines in action," Ph.D. Dissertation, Massachusetts Institute of Technology, Cambridge, MA, 2011.
[208] N. Helms, "MQ-9 Reaper Block 50: Cockpit evaluation report," Cockpit Working Group, U.S. Air Force, Mar. 2013.
[209] J. M. Gilmore, "RQ-4B Global Hawk Block 30: Operational test and evaluation report," DOT&E, May 2011.
[210] S. Ackerman, "The Pentagon doesn't trust its own robots," *Wired Magazine*, 11-Sep-2012. [Online]. Available: http://www.wired.com/dangerroom/2012/09/robot-autonomy/.
[211] *Technology Horizons: A Vision for Air Force Science and Technology 2010-30*. Office of the Chief Scientist of the Air Force, 2011.
[212] *The Role of Autonomy in DoD Systems*. Defense Science Board, 2012.

[213] *Unmanned Systems Integrated Roadmap 2011-2036*. Department of Defense, 2011.
[214] *Unmanned Aircraft Systems Flight Plan 2009-2047*. U.S. Air Force, 2009.
[215] K. H. Abbott, D. McKenney, and P. Railsback, "Operational use of flight path automation systems," Federal Aviation Administration, Sep. 2013.
[216] N. A. Visnevski and M. Castillo-Effen, "A UAS capability description framework: Reactive, adaptive, and cognitive capabilities in robotics," in *IEEE Aerospace Conference*, 2009, pp. 1–7.
[217] N. G. Leveson, N. Dulac, K. Marais, and J. Carroll, "Moving beyond normal accidents and high reliability organizations: a systems approach to safety in complex systems," *Organization Studies*, vol. 30, no. 2–3, pp. 227–249, 2009.
[218] J. H. Saleh, D. E. Hastings, and D. J. Newman, "Flexibility in system design and implications for aerospace systems," *Acta Astronautica*, vol. 53, no. 12, pp. 927–944, 2003.
[219] "Air Force Policy Directive 99-1: Test and Evaluation Process." 1993.
[220] "Edwards Air Force Base Instruction 99-101: 412TW Test Plans." 2013.
[221] "Air Force Instruction 63-501: Acquisition Quality Program." 2009.
[222] D. Tannen, "What's in a frame? Surface evidence for underlying expectations," in *Framing in Discourse*, vol. 14, Oxford University Press, 1993.
[223] P. M. Senge, *The Fifth Discipline: The Art and Practice of the Learning Organization*. Random House LLC, 2006.
[224] "Range Safety Group Standard 319-10: Flight Termination Systems Commonality Standard." Range Commander's Council, Department of Defense, 2010.
[225] R. Coram, *Boyd: The Fighter Pilot Who Changed the Art of War*, Reprint. Back Bay Books, 2004.
[226] J. R. Boyd, "A discourse on winning and losing," U.S. Air Force (Chuck Spinney, Chet Richards, editors), 2010.
[227] D. D. Woods, "Toward a theoretical base for representation design in the computer medium: Ecological perception and aiding human cognition," in *Global Perspectives on the Ecology of Human-Machine Systems*, J. M. Flach, P. A. Hancock, J. E. Caird, and K. J. Vicente, Eds. CRC Press, 1995.
[228] J. Rasmussen, "The role of hierarchical knowledge representation in decisionmaking and system management," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-15, no. 2, pp. 234–243, 1985.
[229] D. Vaughan, *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. Chicago: University Of Chicago Press, 1997.
[230] S. W. A. Dekker, *Just Culture: Balancing Safety and Accountability*. Ashgate Publishing, Ltd., 2012.
[231] J. Henrich, S. J. Heine, and A. Norenzayan, "The weirdest people in the world?," *Behavioral and Brain Sciences*, vol. 33, no. 2–3, pp. 61–83, 2010.
[232] D. A. Wiegmann, H. Zhang, T. L. von Thaden, G. Sharma, and A. M. Gibbons, "Safety Culture: An Integrative Review," *The International Journal of Aviation Psychology*, vol. 14, no. 2, pp. 117–134, Apr. 2004.
[233] E. L. Wiener, B. G. Kanki, and R. L. Helmreich, *Crew Resource Management*. Academic Press, 2010.

[234] "Air Force Instruction 90-1301: Implementing Military Flight Operations Quality Assurance." 2010.
[235] "Army Field Manual 3-22.9: Rifle Marsmanship." 2008.
[236] L. G. Shattuck and D. D. Woods, "Communication of intent in military command and control systems," in *The Human in Command*, C. McCann and R. Pigeau, Eds. Springer US, 2000, pp. 279–291.
[237] "DO-312: Safety, Performance and Interoperability Requirements Document for the In-Trail Procedure in the Oceanic Airspace (ATSA-ITP) Application." Federal Aviation Administration, Mar-2012.
[238] C. H. Fleming, M. Spencer, N. G. Leveson, and C. Wilkinson, "Safety assurance in NextGen," NASA, CR–2012-217553, 2012.
[239] R. C. Crane, "Air Force Flight Test Center Test Plan Preparation Guide." 1999.
[240] "Air Force Instruction 63-131: Modification Management." 2013.
[241] "Edwards Air Force Base Instruction 99-103: 412TW Technical Report Program." 2013.
[242] B. Poulson, "412th Test Wing Author's Guide to Writing Technical Reports." 2014.
[243] "Air Force Instruction 99-103: Capabilities-Based Test and Evaluation." 2013.
[244] "Air Force Test Center Instruction 91-203: AFTC Test Safety Review Policy (Edwards AFB Supplement)." 2015.
[245] "AFMC restructures to cut overhead, make command more efficient," *AFMC News*, Nov-2011. [Online]. Available: http://www.afmc.af.mil/news/story.asp?id=123278315.
[246] "Joint Order 7400.8: Special Use Airspace." Federal Aviation Administration, Feb-2013.
[247] "Joint Order 7610.4: Special Operations." Federal Aviation Administration, Apr-2014.
[248] N. Chung, A. Bhujle, A. Mkrtchyan, and S. Stephen, "Integrating the lean enterprise: 412th Test Wing, Edwards Air Force Base, CA," Massachusetts Institute of Technology, 2012.
[249] N. Chung, "Systems-theoretic process analysis of the Air Force Test Center safety management system," M.S. Thesis, Massachusetts Institute of Technology, Cambridge, MA, 2014.
[250] W. Young and N. G. Leveson, "An integrated approach to safety and security based on systems theory," *Communications of the ACM*, vol. 57, no. 2, pp. 31–35, Feb. 2014.
[251] W. E. Young, "Ph.D. Dissertation," Massachusetts Institute of Technology, Unpublished-2016.
[252] "Air Force Manual 91-223: Aviation Safety Investigations and Reports." 2013.
[253] "Air Force Instruction 91-204: Safety Investigations and Reports (AFMC Supplement)." 2011.
[254] "Air Force Instruction 91-203: Consolidated Occupational Safety." 2014.
[255] "Air Force Instruction 91-202: Mishap Prevention Program (AFMC Supplement)." 2015.
[256] "Air Force Materiel Command Pamphlet 91-2: System Safety Groups." 2007.

[257] "AFTC Initial Test Safety Training (Slides)," Edwards AFB, CA, 2014.
[258] "Department of Defense Instruction 6055.01: Safety and Occupational Health Program." 1998.
[259] "Air Force Policy Directive 90-8: Environment, Safety and Occupational Health Management and Risk Management." 2012.
[260] "Air Force System Safety Handbook." AFSEC, 2000.
[261] "Army Pamphlet 385-16: System Safety Management Guide." 2013.
[262] "Department of Defense Unmanned Systems Safety Guide for Acquisition." OUSD (AT&L), 2007.
[263] "Department of Defense Joint Software System Safety Engineering Handbook." OUSD (AT&L), 2010.
[264] "Range Safety Group Standard 321-10: Common Risk Criteria Standards for National Test Ranges (Supplement)." Range Commander's Council, Department of Defense, 2010.
[265] "Range Safety Group Standard 323-99: Range Safety Criteria for Unmanned Air Vehicles." Range Commander's Council, Department of Defense, 1999.
[266] "Range Safety Group Standard 323-99: Range Safety Criteria for Unmanned Air Vehicles (Supplement)." Range Commander's Council, Department of Defense, 2001.
[267] "Air Force Materiel Command Instruction 99-103: Test Management." 2004.
[268] "Air Force Instruction 62-601: Airworthiness." 2010.
[269] "Air Force Instruction 62-601: Airworthiness (AFMC Supplement)." 2011.
[270] "Air Force Lifecycle Management Center Instruction 62-601: Airworthiness Bulletin - Processes for Delegated Technical Authority." 2013.
[271] "Air Force Test Center Instruction 62-602: Airworthiness." 2014.
[272] "MIL-STD-1530C: Aircraft Structural Integrity Program." Department of Defense, 2005.
[273] "MIL-STD-1798C: Mechanical Equipment and Subsystems Integrity Program." Department of Defense, 2013.
[274] "MIL-HDBK-515: Weapon System Integrity Guide." Department of Defense, 2002.
[275] "Air Force Instruction 63-101: Integrated Lifecycle Management." 2013.
[276] "Air Force Pamphlet 63-128: Guide to Acquisition and Sustainment Lifecycle Management." 2009.
[277] "Air Force Materiel Command Pamphlet 63-101: Lifecycle Risk Management." 2012.
[278] "Air Force Materiel Command Instruction 63-1201: Implementing Operational Safety, Suitability, and Effectiveness and Lifecycle Systems Engineering." 2009.
[279] "MIL-HDBK-514: Operational Safety, Suitability, and Effectiveness for the Aeronautical Enterprise." Department of Defense, 2003.
[280] "Joint Order 7110.65: Air Traffic Control." Federal Aviation Administration, 2012.
[281] "Department of Defense Directive 5030.19: DoD Responsibilities on Federal Aviation." 2013.
[282] "Air Force Instruction 11-202v1: Aircrew Training." 2010.

[283] "Air Force Instruction 11-202v3: General Flight Rules." 2010.
[284] "Air Force Instruction 11-214: Air Operations Rules and Procedures." 2012.
[285] "Air Force Instruction 11-230: Instrument Procedures." 2013.
[286] K. E. Johnson, "Ph.D. Dissertation," Massachusetts Institute of Technology, Unpublished-2016.
[287] "Air Force Instruction 11-2RQ-4v3: Globalhawk Operations and Procedures." 2013.
[288] "Air Force Instruction 11-502v3: Small Unmanned Aircraft Systems Operations." 2012.
[289] R. L. Helmreich, J. R. Klinect, and J. A. Wilhelm, "Models of threat, error, and CRM in flight operations," in *Proceedings of the Tenth International Symposium on Aviation Psychology*, 1999, pp. 677–682.
[290] "Air Force Instruction 51-503: Aerospace Accident Investigations." 2010.
[291] "Department of Defense Instruction 6055.07: Mishap Notification, Investigation, Reporting, and Record Keeping." 2011.
[292] S. A. Shappell and D. A. Wiegmann, "Applying reason: The human factors analysis and classification system (HFACS)," *Human Factors and Aerospace Safety*, vol. 1, no. 1, pp. 59–86, 2001.
[293] R. E. King, "A comprehensive effort to arrive at an optimally reliable human factors taxonomy," in *Proceedings of the 18th International Symposium on Aviation Psychology*, 2015.
[294] "Air Force Policy Directive 90-13: Military Flight Operations Quality Assurance." 2010.
[295] "Air Force Instruction 63-133: Aircraft Information Program." 2010.
[296] "Air Force Instruction 63-140: Aircraft Structural Integrity Program." 2014.
[297] "Air Force Materiel Command Instruction 63-501: Quality Assurance." 2001.
[298] "Air Force Materiel Command Instruction 63-510: Deficiency Reporting." 2006.
[299] "Air Force Technical Order 00-35D-54: Deficiency Reporting, Investigation, and Resolution." 2009.
[300] "Special Publication 800-53: Security and Privacy Controls for Federal Information Systems and Organizations." National Institute of Standards and Technology, 2013.
[301] "List of accidents and incidents involving military aircraft (1975–79)," *Wikipedia*. [Online]. Available: https://en.wikipedia.org/wiki/List_of_accidents_and_incidents_involving_military_aircraft_(1975-79). [Accessed: 05-Aug-2015].
[302] N. T. Spark, *A History of Murphy's Law*. Self Published, 2006.
[303] "Air Force Policy Directive 62-6: Airworthiness." 2010.
[304] D. W. Eidsaune, "Aircraft Accident Investigation Board Report: F-22A, T/N 91-4008," U.S. Air Force.
[305] "HAVE RAIDER Technical Plan," 412TW, Feb. 2015.
[306] "HAVE RAIDER Safety Plan," 412TW, 2015-0800, Feb. 2015.
[307] J. M. Converse and S. Presser, *Survey Questions: Handcrafting the Standardized Questionnaire*, vol. 63. Sage, 1986.
[308] R. M. Pirsig, *Zen and the Art of Motorcycle Maintenance: An Inquiry into Values*. Harper Collins, 2006.