DEFINING, CHARACTERIZING, AND ESTABLISHING "SAFE ENOUGH" RISK THRESHOLDS FOR HUMAN SPACE FLIGHT

by

ROBERT PAUL OCAMPO

B.A., Haverford College, 2003

M.S., Massachusetts Institute of Technology, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Aerospace Engineering Sciences

2016

This thesis entitled:
Defining, Characterizing, and Establishing
"Safe Enough" Risk Thresholds for Human Space Flight
written by Robert Paul Ocampo
has been approved for the Department of Aerospace Engineering Sciences

_____
Dr. David Klaus

_____
Dr. James Nabity

Date_____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline

Ocampo, Robert Paul (Ph.D., Aerospace Engineering Sciences)

Defining, Characterizing, and Establishing "Safe Enough" Risk Thresholds for Human Space Flight

Thesis directed by Professor David M. Klaus

No spacecraft will ever be perfectly safe. Consequently, engineers must strive to design, develop, and operate spacecraft that are safe *enough*. This thesis presents a conceptual framework for defining and characterizing "safe" and distinguishing "safe enough" from "not safe enough." Space Shuttle and Soyuz safety records are presented in the context of this framework, and compared to the safety records of various modes of transportation (automotive, rail, boating, general aviation, commercial aviation) and adventure sport activities (skydiving, mountaineering, SCUBA diving). From these comparisons, a heuristic method for predicting space flight risk is derived. This method, which is built upon the inverse correlation between risk and usage, can coarsely predict risk in the absence of detailed spacecraft data. Based on these predictions, spacecraft risk can either be accepted as "safe enough" or rejected as "not safe enough."

ABOUT THE AUTHOR

Robert Ocampo is an aerospace engineer and scientist. He is also an accomplished mountaineer, pilot, hiker, divemaster, EMT, and mountain rescuer. He has both walked and bicycled across the country (thru-hiking the Appalachian Trail in 2004 and riding his bike from Boston to San Francisco in 2008), and summited over 300 peaks, including all 50 U.S. state highpoints and all 58 peaks above 14,000 feet in Colorado.

This thesis is dedicated to all those who blazed the trail before me:

ad astra per aspera

ACKNOWLEDGEMENTS

The author would like to acknowledge his friends and family for their love, support, and encouragement over the last five years:

To those who provided engineering and academic guidance (as well as funding!), at SNC, the FAA, and CU: thank you. John, Angie, Daniel, Chris, Bru and the rest of the Dream Chasers: you took me under your wings (or were they lifting bodies?) and taught me how to quantify risk. I hope one day I can appropriately qualify my thanks. Henry and Ken, this thesis could not have been done without your support and encouragement. Thank you for taking a leap of faith on an unknown from Colorado (an unknown who, at one point, even had *dreadlocks*!). And to my remarkable committee, including Steve, Roger, and both Jims—you helped me in my quest to add a three-letter suffix to my name. I promise not to become *too* insufferable with the new title (Ocampo, 2016).

Armin, Adam, Mandy, Julie, Becky, Jill, Ally, Alan, Julian, and Dennis: why did you let me write *another* thesis?! I'll forgive you this time, but only because your amazing words of kindness and support have made each of my thesis-writing campaigns bearable.

To my fellow rescuers—notably Chris, Angela, and Tim—and my fellow pilots—in particular, Matt, Javi and Lawrence: you were with me the many times I studied risk from the other side of the spreadsheet. This thesis is all the better because of the hours we spent together, scrambling up mountains, evacing down scree, and soaring through the clouds (under an appropriately-filed IFR flight plan of course).

Page and Jess, Curtis and Bronwyn: thank you for watching over me these last few years. In time, I'll be able to overlook all the kale you forced upon me, but it will be

impossible to overlook the meat and potatoes of your support and friendship—they were (and *are*) far too monumental.

I've studied probabilities as part of my PhD for 5 years now, so Teri, I can say with confidence that you are one in a million. Thank you for supporting me in all the ways that you have.

I would be remiss if I didn't also specifically acknowledge my advisor, David Klaus[1]. Your enthusiasm, knowledge, and passion for space flight are an inspiration, both to me and the many students you've taught. I can only hope to be half the mentor that you are.

Quenton: we did it. One more time on the red line, with no regrets and nothing held back.

Lastly, and once again, most importantly, a very special thank you goes to my wonderfully loving and incredibly supportive parents. You knew I could pull this train over the mountain before I even knew myself. I could not have done this without you. I think I can, I think I can, I think I can…

---

[1] And not just because he's (probably) the only person who will ever actually read this thesis—or my footnotes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACRONYMS

ALT	Approach and Landing Test
ASAP	Aerospace Safety Advisory Panel
BTS	Bureau of Transportation Statistics
CCP	Commercial Crew Program
CMG	Control Moment Gyros
CSA	Canadian Space Agency
CSM	Command Service Module
ESA	European Space Agency
ET	External Tank
EVA	Extravehicular Activity
FAA	Federal Aviation Administration
FARS	Fatality Analysis Reporting System
FHWA	Federal Highway Administration
FMEA	Failure Modes and Effects Analysis
FRA	Federal Railroad Administration
FT	Failure Tolerance
FTA	Fault Tree Analysis
GA	General Aviation
HA	Hazard Analysis
ISS	International Space Station
JAXA	Japan Aerospace Exploration Agency
JSC	Johnson Space Center

LEO	Low Earth Orbit
LM	Lunar Module
LOC	Loss of Crew
LOM	Loss of Mission
LRV	Lunar Roving Vehicle
LV	Launch Vehicle
MMOD	Micrometeoroid and Orbital Debris
NASA	National Aeronautics and Space Administration
NHTSA	National Highway Traffic Safety Administration
NPS	National Park Service
PRA	Probabilistic Risk Assessment
RCS	Reaction Control System
ROS	Russian Orbital Segment
SCUBA	Self Contained Underwater Breathing Apparatus
SLS	Space Launch System
SRB	Solid Rocket Boosters
SRQA	Safety, Reliability, and Quality Assurance
SSME	Space Shuttle Main Engines
STS	Space Transportation System
USCG	United States Coast Guard
USOS	U.S. Orbital Segment

PREFACE

Weeks before he would die in a tragic launch pad fire, astronaut Gus Grissom told a reporter:

> *If we die, we want people to accept it. We're in a risky business, and we hope if anything happens to us, it will not delay the program. The conquest of space is worth the risk of life.*

Today, space flight remains a "risky business." Recent accidents, including the loss of Orbital Science's Cygnus spacecraft, the destruction of Space X's Dragon capsule, and the in-flight death of a Virgin Galactic SpaceShipTwo test pilot serve to underscore this point.

But how risky is too risky? Conversely, how safe is safe enough?

To Grissom and his fellow astronauts, space flight was "worth the risk"—even if the risks were mostly unknown, highly consequential, and poorly quantified. To current and future generations of space explorers, however, this equation may no longer hold true. Therefore, it is worth re-examining the question of "how safe is safe enough" from a fresh perspective, one which blends both objective engineering and logical rationalism. That is the primary goal of this thesis.

The answer cannot bring back those we've lost: the astronauts who died on Apollo 1, *Challenger*, and *Columbia*; and the cosmonauts who perished on Soyuz 1 and Soyuz 11. Nor can it protect those we will lose in the future.

But hopefully the answer can serve as a reminder that—in a very real and very quantitative sense—sometimes the greatest risk in any endeavour can be not risking enough.

Robert Ocampo
United States, Earth
April 12, 2016

CHAPTER 1

BACKGROUND

"After the ship has sunk, everyone knows how she might have been saved."

- Italian Proverb

1.1 Objective

This thesis begins with an historic overview of crewed spacecraft safety, from Vostok to the International Space Station (ISS). This overview is intended to provide insight into the techniques and processes that have historically been used to mitigate space flight risk so that "safe enough" can be achieved.

1.2 Mercury

Project Mercury, America's first human spaceflight program, utilized a single-seat capsule built by the McDonnell Aircraft Company. The capsule was launched on top of a modified tactical missile—the Redstone rocket in the case of early suborbital flights and the Atlas D for later orbital missions. While both missiles had a less than exemplary track record prior to their first manned launches (78% and 54%, respectively—see **Figure 1**, [Cassidy et al., 1964; Swenson et al., 1966]) they were favored for the accelerated Mercury program because of the significant experience base associated with their launch and operations.

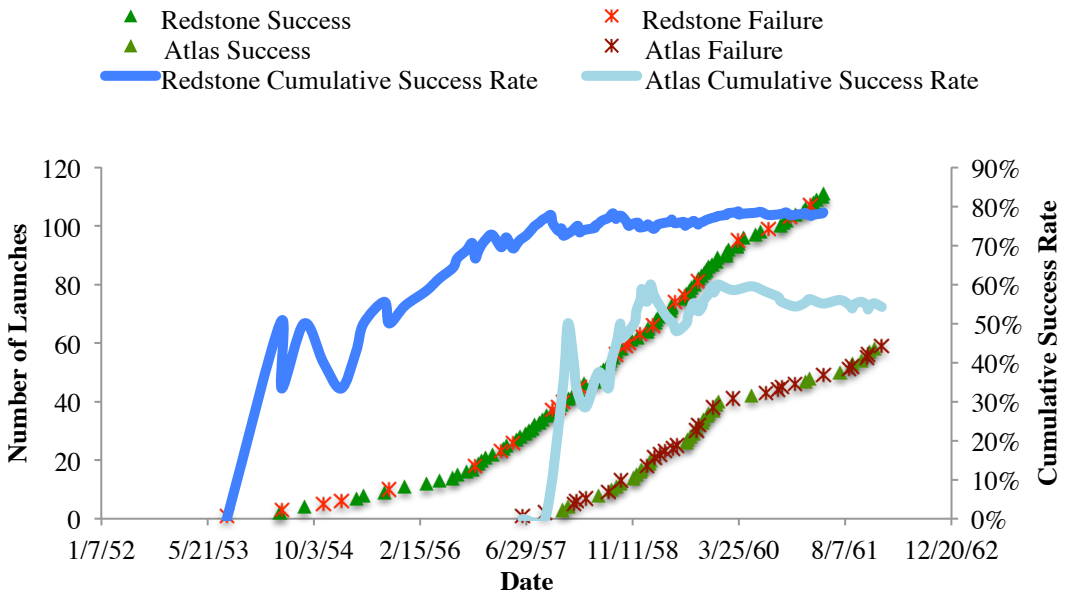


Figure 1: Mercury Redstone and Mercury Atlas success rates and total launches.

Both the Mercury Redstone and Mercury Atlas D shared many broad design characteristics with their uncrewed predecessors. However, both crewed launch vehicles were built to higher quality standards and more conservative design margins. The structure of each rocket, for example, was built to withstand 1.5 times the anticipated loads (Bond, 1988). In addition, both crewed vehicles contained additional redundancy and instrumentation to ensure no single failure could lead to the loss of the mission (French & Bailey, 1963). However, if the crewed rocket *were* to fail catastrophically, an integrated launch escape system was tasked with automatically separating the spacecraft from the launch vehicle.

Risk was further mitigated through extensive ground and flight testing. Hardware was tested iteratively—first at the component level, then as a completed subsystem, and finally as an integrated vehicle (Burkhalter & Sharpe, 1990). Components that could not

be adequately tested on the ground, such as the ablative heat shield or the launch escape system, were tested in flight using the Little Joe or Big Joe boosters (Swenson et al., 1966). As a final precaution, both Mercury Redstone and Mercury Atlas boosters were flown in an unmanned configuration several times prior to their first manned launch.

Organizational procedures also served to improve astronaut safety. Spacecraft and launch vehicle were built with parts identified by a "Mercury stamp," thereby ensuring only qualified components were used in the vehicle (Burkhalter & Sharpe, 1990). Workers were actively encouraged to meet high standards of workmanship, as those that met certain high performance criteria were awarded with marks of distinction. As further incentive, Mercury astronauts made a point of visiting NASA contractors so workers would associate a "face" with the vehicle they were building (Swenson et al., 1966).

Despite the effort made to improve both booster and capsule reliability, each manned Mercury launch suffered its share of hardware failures. In many of these situations, the astronaut successfully served as a final line of defense against mission failure. Originally, the Mercury spacecraft was intended to be fully-automated; the astronaut would fly as a passenger, not as a pilot. However, the astronauts strongly objected to this "spam-in-a-can" design, and a small viewport and manual control system were added to the spacecraft. This allowed the human astronaut to serve as a backup to the automated flight control. This design choice proved particularly effective during the last manned Mercury mission, allowing Gordon Cooper to pilot his Faith 7 spacecraft through reentry after his automatic stabilization and control systems were lost (Swenson et al., 1966).

The human-rating process (see sidebar below) for Project Mercury proved to be a significant challenge, both in terms of schedule and cost: With roughly 80,000 critical parts in the capsule and booster, the first manned Mercury Redstone launch took place over a year behind schedule and cost 40% more than its unmanned predecessor (Swenson et al., 1966). Despite these modifications, the reliability of the Redstone only increased from 81% (the success rate of the rocket prior to 1961) to 84% (the reliability estimate of the Mercury Redstone rocket) (Cassidy et al., 1964). Ultimately however, the human-rating process for Mercury proved effective, as all 6 astronauts returned safely from their Mercury flights.

SIDEBAR: What is human-rating?

Human-rating (or its functionally equivalent precursor, "man-rating") is a phrase that originated in the mid-20th century to describe hardware developed specifically for manned use or occupation. The first vehicles to be human-rated were the X-series of experimental rocket planes (Bond, 1988; Heppenheimer, 2002). For this reason, the term "human-rating" is most commonly associated with aircraft and spacecraft.

Originally, human-rating focused predominately on crew safety. The Redstone, Atlas, and Titan II rockets—the military missiles adapted for the U.S. Mercury and Gemini programs—had a success rate of only 81%, 75% and 74% prior to their first crewed flights (Cassidy et al., 1964; Swenson et al., 1966). To improve the likelihood of crew survival and mission success, NASA began using off-the-shelf components (to improve subsystem reliability), added redundancy (*continued*)

(*continued from previous page*) to critical systems, and developed a launch escape system (Swenson et al., 1966; Bond, 2002; French & Bailey, 1963).

As Mercury and Gemini evolved into Apollo and Skylab, human-rating began to focus on improvements in operability as well as safety. As noted in a 1988 NASA document, "the human rating process for the Mercury, Gemini, and Apollo Programs was centered on human safety. The Skylab and Shuttle Programs added to this an emphasis on human performance and health management" (Zupp, 1995, p. 1).

Despite these additions, human-rating remained a rather generic concept during the 1970s and 1980s—it was applied to *any* system that could transport and/or support humans in space (Musgrave et al., 2009), rather than to a specific type of vehicle. A set of guidelines from Johnson Space Center (JSC) attempted to bring clarity to the term in 1988 by defining a human-rated system as one that required an escape system or safe haven (NASA, 1988). Based on this definition, the Space Shuttle was not considered by the JSC group to be human-rated but rather "Highly Reliable."

It wasn't until 1992 that human-rating began to take its current shape as a requirements-based methodology. That year, NASA formed a committee to develop a set of human-rating requirements (Zupp, 1995), which eventually evolved into JSC 28354 and ultimately NASA NPR 8705.2, the agency's "Human-Rating Requirements for Space Systems." This document defines a human-rated system as one that "accommodates human needs, effectively utilizes human capabilities, (*continued*)

(*continued from previous page*) controls hazards and manages safety risk associated with human spaceflight, and provides, to the maximum extent practical, the capability to safely recover the crew from hazardous situations." (NASA, 2008a, p. 4)

1.3 Gemini

Gemini was intended to bridge the gap between Mercury and Apollo, with missions designed to parse out the techniques and technologies required for rendezvous, docking, long-duration flight, and extra-vehicular activity (EVA). McDonnell Aircraft was once again selected to build the two-person spacecraft, which was launched on a modified Titan II intercontinental ballistic missile. Later missions incorporated the use of an Agena upper-stage booster, which served as a docking target and third-stage booster for the Gemini spacecraft.

Like the Atlas and Redstone rockets before it, Titan II was originally designed for military applications, then later adapted for human use. These modifications included the addition of redundant hydraulic, electrical, and flight control systems; an upgraded factor of safety for structural components (1.25); and the inclusion of a malfunction detection system (Bond, 1988). Oxidizer standpipes and mechanical accumulators were also added to the booster to eliminate longitudinal "Pogo" oscillations that often occurred during launch (Hacker & Grimwood, 1977).

Prior to its first crewed launch, the Titan II booster had accrued a significantly higher success rate than either Mercury Redstone or Mercury Atlas (see **Figure 2**), and benefited from concurrent reliability improvements initiated by the crewed Dyna-Soar-Titan II program. Moreover, Titan II boosters assigned for crewed use were built in a

facility separate from other missile production lines to further improve quality (Franzini & Fragola, 2011).

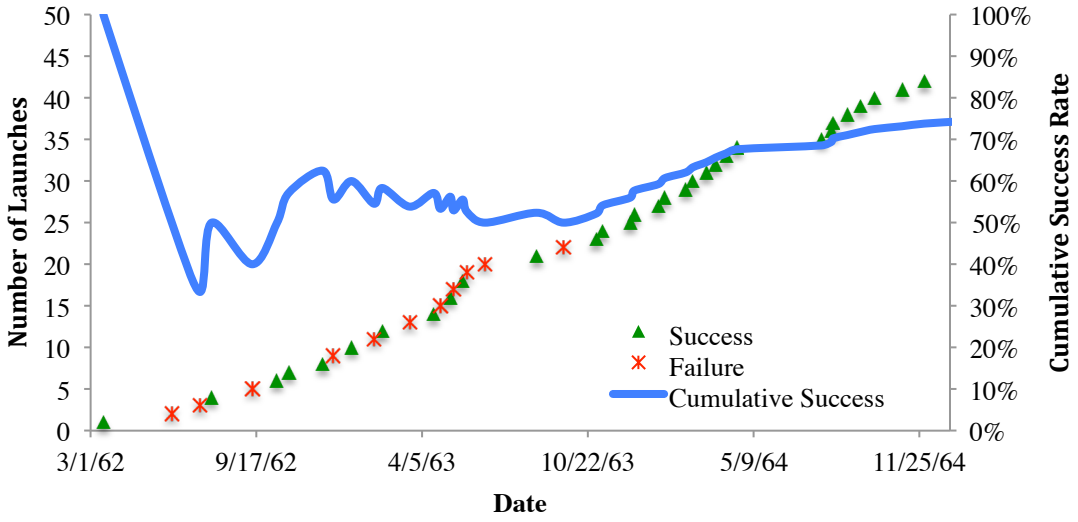


Figure 2: Number and cumulative success rate of Titan II launches before first crewed Gemini flight.

The Gemini spacecraft inherited a number of flight-proven subsystems from its Mercury predecessor. "Lessons learned" during Mercury capsule design and construction were captured and faithfully passed down to Gemini engineers (a process aided by the fact that the same contractor, McDonnell Aircraft, built both vehicles). However, the *location* of these subsystems differed substantially in the newer spacecraft. Due to the thrust limitations of the Mercury launch vehicle, the Mercury capsule incorporated integrated systems, attached in the manner of a "layer cake." While this technique significantly decreased mass, it made spacecraft testing and checkout burdensome. In contrast, the Gemini spacecraft utilized a separate "service module"

containing modularized subsystems, a design which significantly expedited and improved verification and checkout (Hacker & Grimwood, 1977).

Unlike its programmatic predecessor, Gemini lacked an escape tower. Instead, the capsule incorporated ejection seats designed to separate the crew from the spacecraft during a launch and landing emergencies. This abort system methodology was chosen ostensibly to simplify and "modularize" the design, but proved difficult to implement in practice (a malfunction during testing destroyed a test dummy) (Hacker & Grimwood, 1977). Notably, ejection could only be initiated manually, a technique in line with the greater flight control authorities allotted to astronauts during Gemini (Embrey, 1966) and very much appropriate given the Titan II's hypergolic propellants. The decision to incorporate manual ejection capabilities proved well-founded when a tower plug prematurely separated from the Gemini 6-Titan II rocket prior to liftoff. Although mission rules called for an ejection, the astronauts (appropriately) elected to remain in their spacecraft, thereby salvaging the mission (Hacker & Grimwood, 1977).

Originally intended as an add-on to Mercury, Gemini suffered from significant cost overruns as it developed into its own full-fledged, stand-alone program. Because of budget constraints and schedule pressures, Titan II engine test firings were curtailed and quality assurance and reliability testing programs were eliminated, replaced instead with cheaper enhanced qualification testing. The effects of such a fast-paced program were not inconsequential: Thrusters aboard Gemini 7 failed towards the end of flight because those installed were of an older design known to have problems (Hacker & Grimwood, 1977).

Although all 10 Gemini missions ended with the crews' safe return, Gemini 8 nearly ended in catastrophe. Upon docking with its Atlas Agena target, a stuck thruster in the spacecraft began rolling the spacecraft at a rate that threatened to cause the crew to lose consciousness. After manually shutting down the thruster and activating the reentry control system, the crew was able to stabilize their spacecraft and initiate an emergency landing in the Pacific Ocean (Hacker & Grimwood, 1977).

1.4 Apollo

The Apollo program safely landed 12 men on the moon between 1969 and 1972. The 3-man crew utilized two separate spacecraft on their lunar missions: the Command and Service Module (CSM), which served as primary crew quarters and Earth-entry vehicle, and the Lunar Module (LM), which provided two astronauts with lunar landing and ascent capabilities. Both the CSM and LM were launched on the Saturn series of vehicles. Saturn IB rockets were utilized for low-earth orbit missions; Saturn V rockets were used primarily for lunar voyages.

Unlike boosters used in Mercury and Gemini, the Apollo Saturn rocket was designed *explicitly* for human use (Harris & Brom, 1965). Human-rating features were built into the vehicle from the start (rather than being grafted on later), with redundant systems eliminating most single point failures. Moreover, the vehicle's design was inherently conservative: The Saturn series of rockets relied on state of the art (*not* cutting edge) technologies and margins that were considered "lavish even by aerospace standards" (Murray & Cox, 1989, p. 38). And, if the booster *were* to fail catastrophically,

an emergency detection system and abort tower were available to rapidly separate the spacecraft from the launch vehicle (Embrey, 1966).

The nascent Saturn rockets had an attendant disadvantage, however: a knowledge base for the rocket did not exist prior to Apollo (Harris & Brom, 1965). To validate the Saturn's design while maintaining the pace necessary to meet President Kennedy's lunar landing goal, engineers employed a technique known as "all up testing" in which *all* stages of the vehicle were flown live on each launch. In this manner, a successful test of the lower stages could provide flight data for the upper stages. This technique largely contributed to Saturn's accelerated human-rating process (Bilstein, 1980).

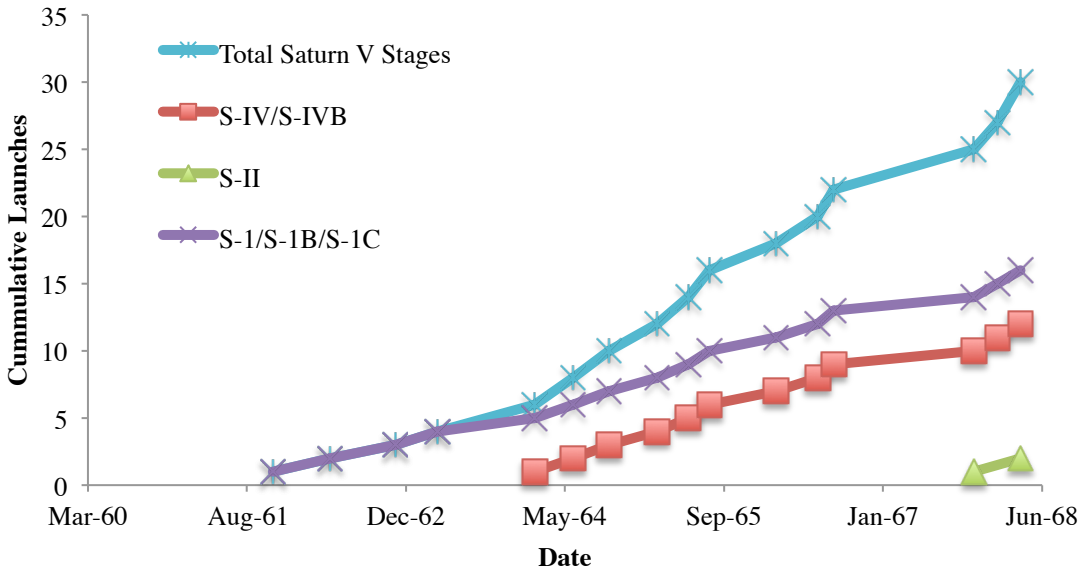


Figure 3: Total number of launches by Saturn stage.

The rocket's human-rating was also aided by Saturn's modular design. Because many stages were interchangeable (the S-IV earth departure stage, in some derivation,

appeared on the Saturn 1, the Saturn 1B, and the Saturn V), data accrued during early uncrewed Saturn 1 and Saturn 1B launches could be applied to later crewed launches of the Saturn V (see **Figure 3**). Given the S-IV's early and frequent success, NASA felt confident launching humans to the moon on the very first manned Saturn V (Bilstein, 1980).

Once in orbit, the crew traveled to and from the moon in the CSM and LM. Despite their inherent complexity—the combined CSM/LM had over 3 million parts (Bilstein, 1980)—both spacecraft were designed to extremely high standards of reliability. North American Aviation, charged with designing the Command Service Module, utilized proven technologies and employed redundant components wherever possible. The Lunar Module, built by Northrop Grumman, aimed for reliability through simplicity (Brooks et al., 1979); the fixed ascent engine on the LM, for example, utilized a pressure-fed engine hypergolic fuel and oxidizer, thereby negating the need for an igniter (and thus removing a potential failure mode) (Brooks et al., 1979). Even the Lunar Roving Vehicle (LRV), utilized in later Apollo missions to extend the astronaut's travel range, adhered to strict human-rating requirements. Through design and operations, the LRV was single-fault tolerant to Loss of Mission (LOM) and dual-fault tolerant to Loss of Crew (LOC) (Young, 2007).

Although Apollo successfully met President Kennedy's goal of landing men on the moon and safely returning them to Earth before 1970, the program was not without its share of failures. In 1967, a fire in the command module during a "plugs-out" test claimed the lives of astronauts Gus Grissom, Ed White, and Roger Chaffee. A frayed wire beneath the command module pilot's seat is thought to have triggered a spark, and

the CSM's high pressure, 100% oxygen crew environment—coupled with an abundance of flammable materials in the cabin—contributed to the fire's rapid, lethal spread (Apollo 204 Review Board, 1967). A second failure of the CSM—this time involving a high pressure oxygen tank—nearly claimed the lives of a second crew three years later when an oxygen tank in the Apollo 13 service module exploded halfway to the moon, forcing the crew to retreat to their lunar module. The LM, though not designed for such a contingency, successfully served as a "lifeboat" and the crew returned to Earth safely (Apollo 13 Review Board, 1970).

1.5 Skylab

The Skylab space station, launched in 1973, hosted three separate American crews over the course of a nine month period. During 28, 59, and 84 day missions, Skylab astronauts conducted experiments in astronomy, physiology, biology, and remote sensing. Leftover Saturn hardware served as both the station's backbone and its transportation infrastructure: a modified Saturn S-IVB stage, boosted by an unmanned Saturn V rocket, functioned as the station's orbital workshop and crew quarters, and an Apollo Command and Service Module (CSM), launched on a Saturn IB booster, provided crew transportation to and from the station.

During its launch to orbit, Skylab suffered critical damage to its electrical and thermal protection systems. A micrometeoroid shield, used to both protect and cool the station, broke loose, knocking out one of two primary solar arrays. Initially engineers feared that such damage was beyond repair; however, by deploying a temporary "parasol" and manually deploying the station's remaining solar array, Skylab astronauts

were able to restore the station to near-nominal functionality. A more permanent sunshade—the "Marshall sail"—was subsequently installed by the 2nd Skylab crew. In-flight maintenance and operational procedures mitigated the effects of later coolant system leaks and Control Moment Gyro (CMG) failures (Hitt et al., 2008).

Designed to support crews of astronauts for upwards of a year, Skylab was subject to numerous human-rating requirements. Only parts that had already been proven in space or rigorously tested on the ground could be used on board the station. Moreover, NASA limited its selection of Skylab contractors to those that had successfully flown flight hardware in the past. As a final safeguard, components that were deemed critical were designed as single-fault tolerant or exceptionally reliable (Belew & Stuhlinger, 1973).

All 3 Skylab crews completed their missions and returned to Earth safely. However, several hardware failures on board the Apollo spacecraft threatened to curtail two of the missions. The first Skylab crew was forced to initiate a "hard dock" maneuver to link their spacecraft to the space station when capture latches on the CSM port failed to engage (Hitt et al., 2008). During the second manned Skylab mission, two of the four Reaction Control System (RCS) jets on the Service Module failed in orbit, threatening to strand the crew in space. A potential rescue mission was initiated but never launched, as the crew managed to deorbit their spacecraft with the remaining RCS jets (Hitt et al., 2008).

1.6 Space Shuttle

From 1981-2011, the US Space Shuttle—the world's first partially reusable spacecraft—performed a variety of missions in Low Earth Orbit (LEO). Over the course of 135 flights, shuttle crews deployed and retrieved satellites, performed experiments in Spacelab and SPACEHAB scientific modules, resupplied the Soviet *Mir* space station, and helped assemble the International Space Station (ISS).

Launched in a multi-stage, parallel-burn configuration, three Space Shuttle Main Engines (SSMEs), fueled by an External Tank (ET) and augmented by twin Solid Rocket Boosters (SRBs) provided thrust to the crewed orbiter during ascent. During landing, the winged orbiter returned to Earth as an unpowered glider, landing on a runway (Stockton & Wilford, 1981).

Given the diversity of its mission objectives and the complexity of its flight operations, shuttle development proved extremely challenging. Building a reusable spacecraft necessitated major advances in thermal protection, computer avionics, and propulsive engineering (Heppenheimer, 2002). The Space Shuttle Main Engines "required a greater step forward in technology over the Saturn engines used in Apollo than the Saturn engines did over their predecessors" (Stockton & Wilford, 1981, p. 56). Yet despite the vehicle's heavy reliance on unproven technologies, the space shuttle was never tested in an uncrewed configuration; both its first Approach and Landing Test (ALT) *and* its first launch were crewed. To certify the shuttle as safe for flight, NASA relied solely on ground testing in conjunction with model analysis (CAIB, 2003).

If critical components were to break down in flight, redundant spares provided fault tolerance (Williamson, 1999); if engines were to fail during launch, several abort

modes were available. As a last resort, the crews of the first four "developmental" flights had the option of ejecting if a catastrophic malfunction were to occur. In 1988 (after the *Challenger* disaster), a sliding pole escape system was added to the orbiter to allow for crew bailout during certain phases of compromised launch and landing operations.

The Space Shuttle was the only NASA program to lose crew members in flight. In 1986, the orbiter *Challenger* broke apart 73 seconds after launch. Heated gas from an SRB field joint breached both primary and secondary O-ring seals, impinging upon and destroying the ET-SRB attachment strut. This event led to the aerodynamic destruction of the vehicle and loss of the entire crew (Rogers et al., 1986).

17 years later, the Orbiter *Columbia* disintegrated during re-entry, killing all 7 crewmembers on board. Insulating foam from the External Tank broke loose during launch, colliding with and damaging the thermal protection system on the shuttle wing leading edge. During re-entry, heated plasma breached the affected wing, melting the spacecraft's aluminum structure and destroying the vehicle (CAIB, 2003).

Both accidents were presaged by anomalies that indicated serious weaknesses in the shuttle system: O-ring "blow-by" occurred 10 times prior to *Challenger*; ET foam shedding was identified 6 times prior to *Columbia* (Rogers et al., 1986; CAIB, 2003). The Rogers Commission and the Columbia Accident Investigation Board (CAIB)—the investigatory boards formed in the wake of the two shuttle accidents—asserted that engineers had disregarded these anomalies in the face of budget and schedule pressures (CAIB, 2003; Rogers et al., 1986). NASA responded by modifying shuttle hardware, upgrading safety standards, and revamping its Safety, Reliability, and Quality Assurance (SRQA) programs.

1.7 International Space Station (ISS)

The International Space Station (ISS) is a modular space laboratory designed and built by the United States, Russia, Japan, Canada, and partner nations from the European Space Agency (ESA). The first ISS module was launched in 1998; after extensive delays following the space shuttle *Columbia* disaster, the station was completed in 2011.

Although structurally unified, ISS is *programmatically* divided into Russian and U.S. Orbital Segments (ROS and USOS, respectively, with ESA, CSA, and JAXA hardware being considered part of the USOS). Such segmentation offers dissimilar failure tolerance to critical and catastrophic hazards (ISS Independent Safety Task Force, 2007). If all four U.S. Control Moment Gyros (CMGs) were to fail, for example, (as one did in 2002 and again in 2006) thrusters on the Russian Service Module can provide backup attitude control. The benefits of segmentation, however, come at a price: hardware built in one country must integrate cohesively and safely with hardware created elsewhere—a significant challenge given that *system-wide* testing and verification of the ISS was not accomplished prior to the start of ISS on-orbit assembly (ISS Independent Safety Task Force, 2007).

During its 17 years in orbit (as of November, 2015), the ISS has suffered several critical component failures (ISS Independent Safety Task Force, 2007). In 2004, the *Elektron* oxygen generator broke down, forcing the crew to rely on Solid-Fuel Oxygen Generator (SFOG) "candles" for oxygen—the very same candles responsible for the fire on *Mir*. Two years later, a similar *Elektron* unit began leaking potassium hydroxide, a toxic irritant; although the situation was eventually stabilized, the crew on board was

obliged to don masks and surgical gloves as a precautionary measure until the atmosphere was cleared.

External hazards, such as Micrometeoroid and Orbital Debris (MMOD), have also posed threats to the ISS. In 2009 and 2011, large pieces of debris nearly collided with the station; and in 2012, a small MMOD object actually struck (but did not penetrate) a window on the ISS cupola. Although the ISS design was intended to meet a 95% probability of no penetration of pressurized compartments, certain Russian segments, originally designed for the Russian Mir2 station, were not designed to this same standard (ISS Independent Safety Task Force, 2007).

Despite the criticality of these incidents, according to ESA, station-wide safety procedures remain underdeveloped (Pelton & Marshall, 2006). There remains no unified ISS Safety Authority, and political sensitivities continue to limit international information transfer. Nevertheless, at this time the United States expects to support the USOS segment of the ISS until at least 2024, while Russia hopes to eventually utilize their segment as the building block of a third-generation space station (Zak, 2009).

1.8 Comparing the U.S. and Soviet/Russian Space Programs

Although the technical aspects of spaceflight remain the same whether one launches from Baikonur or Cape Canaveral, significant *philosophical* differences separate the Soviet/Russian and U.S. space programs. These differences are driven in large part by programmatic and socio-political influences (Hall & Shayler, 2001; Chertok, 2009; Ivanovich, 2008; Gibbons, 2008; Shelton 1968; Hall & Shayler, 2003; Harland, 2007).

• Historically, the Soviet/Russian space program has been less open to the public and more accepting of risk than its US counterpart.

• The Soviet/Russian space program has approached spacecraft design from an evolutionary, rather than a revolutionary, perspective—the current Soyuz spacecraft and Soyuz rocket are part of an engineering lineage that stretches back 40+ years.

• Having more experience with long-duration spaceflight than the US, the Soviets/Russians are accustomed to relying on repair as a means of ensuring spacecraft reliability.

• The Soviet/Russian program assigns less autonomy to their cosmonauts, relying instead on flight controllers on the ground and/or automated systems on the spacecraft for critical decisions and actions. In contrast, the United States typically allows considerably more crew control of spacecraft and launch vehicle functions.

Despite these differences, the Soviet/Russian and US programs have comparable flight safety records, with each having lost only 2 crews in 50+ years of spaceflight. The techniques and processes used to reduce risk on Soviet/Russian space flights are described below.

1.9 Vostok/Voskhod

The Soviet Vostok program succeeded in launching the first human space flight, the first-multi-orbit and multi-day missions, and the first set of tandem spaceflights. The single-seat capsule (Vostok 3KA) was launched on a variant of the R-7 Inter-Continental Ballistic Missile (ICBM) known as the Vostok-K (8K72K). Like its American counterpart, the Mercury-Atlas, the R-7 had a relatively poor track record prior to its first manned launch, succeeding only 57% of the time (see **Figure 4**). (According to Hall et al., 2001 (p. 56), it was Soviet practice to carry out "more flight-testing than trouble shooting before flight tests"; this may in part explain the R-7s relatively low early success rate. Nevertheless, most Soviet engineers considered the launch vehicle to be the weakest link of the Vostok program [Chertok, 2009]). As such, ejection seats, which were nominally used during landing, were also made available for ascent emergencies.

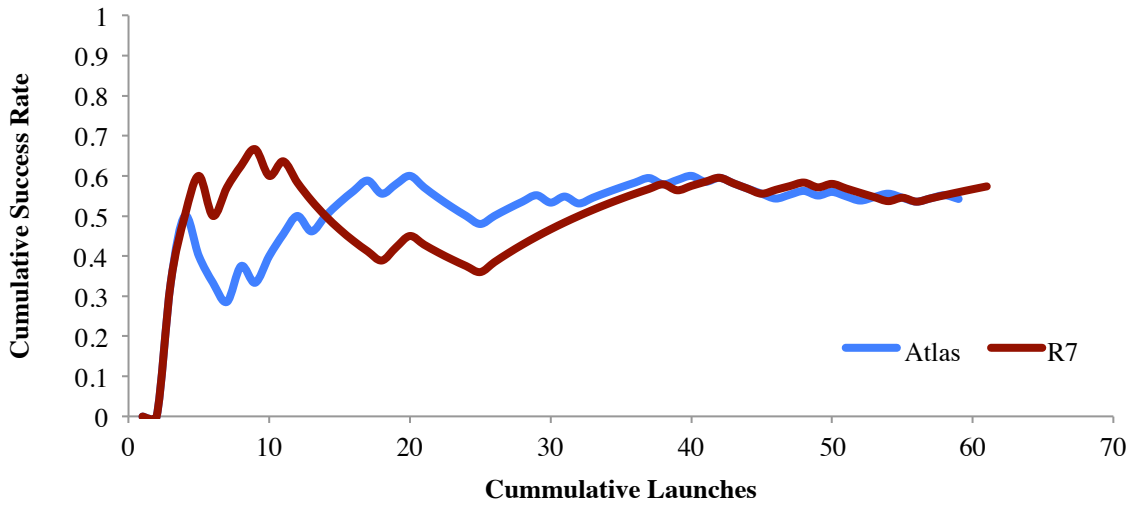


Figure 4: R-7 and Atlas cumulative success rates prior to their first crewed launch.

To improve the reliability of the vehicle during flight, a strict quality control and testing program was put in place for Vostok. Every aspect of the spacecraft's fabrication underwent "painstaking examination" and a "complete cycle of factory tests" before being delivered to the launch site (Chertok, 2009, p. 52 & p. 20). Parts that passed inspection were then logged as "suitable for 3KA" to differentiate them from unmanned R-7 missile components (a technique analogous to one used in Project Mercury).

Functional redundancy and design margins also served to improve spacecraft safety. The spacecraft's pressurization and control systems were designed to withstand a single fault (Chertok, 2009), and life support consumables were sized to last until the natural decay of the vehicle's orbit (thereby mitigating the effects of spacecraft retrorocket failure—a very real risk given its occurrence on the unmanned Korabl-Sputnik 1). Notably, Vostok differed from the Mercury capsule in that manual control *did not* serve as a means of redundancy.

In 1964, Vostok was succeeded by Voskhod, an upgraded multi-crewed capsule with redundant re-entry rockets, an added descent braking engine, and in one instance, an EVA airlock. In order to accommodate multiple crew members, Voskhod cosmonauts were launched without ejection seats, abort tower, or pressure suits.

Both Voskhod flights and all six Vostok flights ended with the cosmonauts' safe return; however, several close-calls occurred during re-entry. On Vostok 1, 2, 5, and Voskhod 2, the instrument module failed to disconnect from the descent module, causing the spacecraft to tumble until the dynamic pressure of re-entry could separate the two segments. Voskhod 2 also suffered from a failure of its automated re-entry system, forcing the two cosmonauts to rely on their backup manual re-entry system. The

spacecraft landed several thousand miles off course, and the cosmonauts were not recovered until 48 hours after landing.

1.10 Soyuz

The Soyuz spacecraft has been the mainstay of the Soviet/Russian manned space program. First launched in 1967, Soyuz has supported 129 crews on 7 different spacecraft variants (**Table 1** shows a summary of Soyuz spacecraft as of March 2016). Although it was originally designed for the Soviet manned lunar program and actually flew several unmanned Zond circumlunar flights (Chertok, 2009), Soyuz has since proven its merit as a space station transfer vehicle, shuttling crews to Salyut, Mir, and the International Space Station. (During the 1970s, six Soyuz missions ended prematurely due to rendezvous or docking failures; however, in the intervening years, Soyuz has since improved its track record (Hall & Shayler, 2003).

The Soyuz spacecraft is launched on top of the Soyuz booster, a derivative of the R-7 ICBM. Throughout the years, this launch vehicle has proven exceedingly reliable—with 700+ launches to its credit, the Soyuz booster maintains a success rate that exceeds 97%. Despite this exceptional track record, all manned Soyuz missions are launched with an automatic launch escape system; additionally, all Soyuz subsystems are designed to be one fault tolerant to loss of mission, and two fault tolerant to loss of crew (Chertok, 2009). As a final precaution, all spacecraft systems undergo thorough testing prior to flight (Hall & Shayler, 2003).

Soyuz Variant	Year(s)	Launches
Soyuz 7K-OK/OKS	1967-1971	10
Soyuz 7K-T	1973-1981	26
Soyuz 7K-TM	1975	3
Soyuz-T	1976-1986	15
Soyuz-TM	1986-2002	33
Soyuz-TMA	2003-2012	22
Soyuz-TMA-M	2010-present	20

Table 1: Soyuz variants, launch dates, and number of launches, as of March, 2016.

During the last four decades, the Soyuz spacecraft has undergone a series of modifications aimed at incrementally improving cost, safety, and mission assurance. However, these changes have been evolutionary, rather than revolutionary in nature; as such, the current design retains (and benefits from) both state of the art hardware *and* flight-proven subsystems. (Many Soyuz components were previously or concurrently incorporated on Kosmos, Zond, Progress, and Salyut spacecraft [Chertok, 2009]).

Nevertheless, Soyuz has suffered its share of critical and catastrophic failures, primarily in the early years of its history. In 1975, the Soyuz 18a booster failed to stage, leading to the automated separation of its capsule prior to orbital insertion. Eight years later, cosmonauts aboard Soyuz T-10a were the first to survive a pad abort after their Soyuz booster caught fire on the launch pad.

Critical and catastrophic incidents have also occurred during re-entry and landing. Cosmonauts on Soyuz 23 landed in a freezing lake, and were rescued only a few hours

before their consumables were depleted. In 1967, cosmonaut Vladimir Komarov perished when his parachute failed to deploy on Soyuz 1. Four years later, three cosmonauts died when a pressurization valve aboard their Soyuz 11 spacecraft inadvertently opened during re-entry. Both catastrophic incidents have been attributed to a flawed quality control system (Chertok, 2009).

1.11 Salyut

In 1971, the Soviet Union launched Salyut 1, the world's first space station. In the decade to follow, the original Salyut was succeeded by six 1st generation and two 2nd generation Salyut stations. Of these nine space stations, three were destroyed during launch or in the early days of its mission (Ivanovich, 2008).

Due to Salyut's close ties with the military Almaz space station, many details regarding Salyut hardware remain classified. However, evidence suggests that a number of subsystems used in Salyut were first flight-tested in the manned Soyuz and unmanned Zond programs (Gibbons, 2008).

No cosmonauts were lost while *aboard* Salyut space stations (although the crew of Soyuz 11 did perish after leaving Salyut 1); however, several critical incidents occurred, including a small electrical fire aboard Salyut 1, a (potential) ECLSS failure on Salyut 5, and a fuel leak on Salyut 7 (Ivanovich, 2008). In 1985, a cosmonaut, Vladimir Vasyutin, was evacuated from the station prior to the completion of his mission. Additionally, six missions to Salyut stations were curtailed by rendezvous or docking failures (see **Figure 5**).

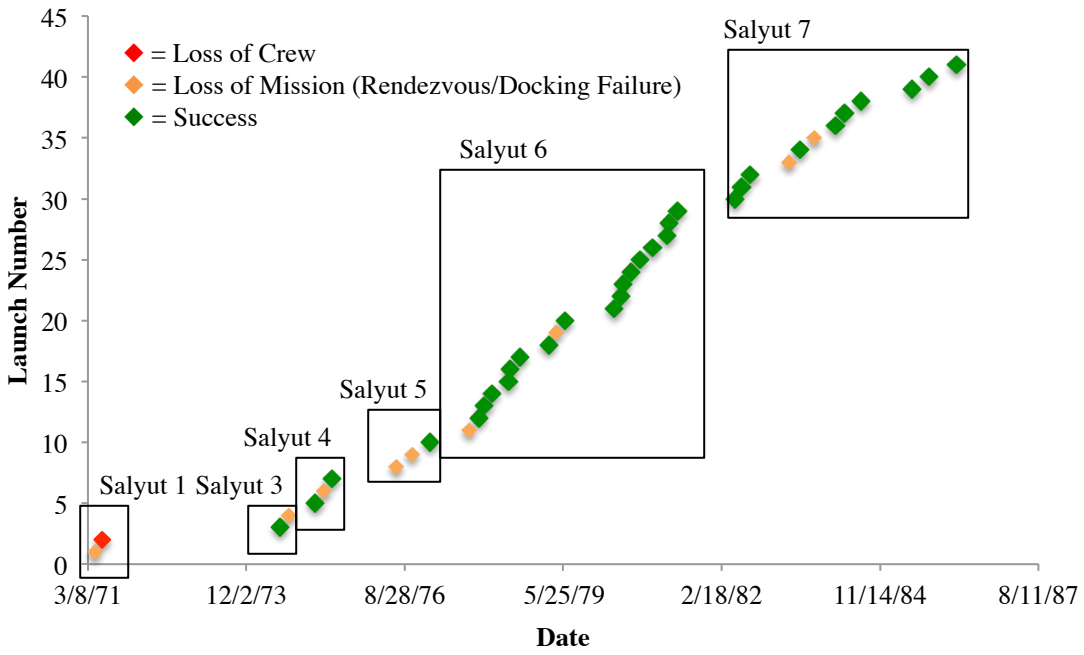


Figure 5: Salyut mission success over time. It should be noted that the Loss of Crew event that occurred during Salyut 1 was not due to the Salyut station itself, but rather the Soyuz return vehicle.

1.12 Mir

The Soviet (later, Russian) Mir was the first space station to be assembled on orbit in piecemeal fashion. The first module, the base block, was launched in 1986; six additional modules were added in the decade that followed. Presaging the docking mishaps that would plague Mir in the 1990s, the first three modules to be added—*Kvant 1*, *Kvant 2*, and *Kristall*—all suffered from initial automated docking failures before successful re-rendezvous and attachment (Harland, 2007).

During its 15 years in orbit, Mir greatly exceeded its design lifetime, in some cases by over a decade. Yet despite Mir's longevity, subsystem failures proved constant

during its later years of operations, particularly with respect to the life support and thermal control systems (Burrough, 1998). Redundancy, resupply, and crew maintenance succeeded in mitigating the effects of many of these failures. **Table 2** shows a summary of critical mishaps and failures onboard Mir.

Type	Category	Mission	Year	Event
Collisions	Near-Misses	Progress M-7	1991	Passes within 5m of station
		Progress M-33	1997	Passes within 10m of station
	Incidental	Soyuz TM-17	1993	Collides with Kristall
		Progress M-24	1994	Collides with Mir
	Critical	Progress M-34	1997	Causes depressurization of Spektr
Fire	Incidental	Mir EO-17	1994	Vika oxygen fire
		Mir EO-23	1997	SFOG oxygen fire
Medevac	Critical	Soyuz T-14	1985	Medevac due to crew illness

Table 2: Near-Miss, critical, and incidental events on Mir.

Mir suffered from several critical and near-catastrophic fires. In 1994, a fire in the *Vika* oxygen-producing systems broke out on Mir, but was smothered before it could spread. Three years later, another oxygen fire started in *Kvant 1*. Although the crew extinguished the fire before it could engender catastrophe, the fire severely charred the walls of the module and generated significant levels of toxic smoke (Burrough, 1998; Linenger, 2000).

Mir also suffered a number of collisions and "near-misses" with manned Soyuz transfer ferries and unmanned Progress freighters (Mir suffered much smaller collisions as well, namely in the form of Micrometeoroids and Orbital Debris). In 1994, Mir passed through the remains of the Swift –Tuttle comet and was impacted over 60 times [Harland, 2007]). Progress M-7 and Progress M-33 narrowly avoided collision with Mir when automated control was lost during final approach. Progress M-24 and Soyuz TM-17

actually struck the station, but did not cause life-threatening damage. In 1997, Progress M-34 collided with *Spektr* during a test of the manual docking system, causing depressurization of the module. Only by sealing *Spektr* from the remaining habitat modules was the crew able to avert disaster.

Mir was deorbited in 2001, after being visited by 39 crews from 12 countries. Modules for the follow on Mir 2 were eventually utilized on the Russian segment of the International Space Station (Bond, 2002).

1.13 Chapter Summary

The techniques and processes described here were (and *are*) intended to make space flight safe as possible. Of the 304 manned missions launched by the governments of the United States and Soviet Union/Russia between 1961 and 2016, only 4 have resulted in catastrophic (i.e. fatal) in-flight accidents. This amounts to a success rate of 98.7%.

1.14 Immediate Forward Work

Identifying a process that can determine whether this success rate—or *any* success rate for that matter—is "safe enough" is the primary goal of this thesis, and will be discussed in detail in the chapters that follow.

1.15 Related Publications, Presentations, and Posters

Klaus, D., Fanchiang, C., & Ocampo, R. (2012). Perspectives on Spacecraft Human-Rating. In *42nd International Conference on Environmental Systems* (p. 3419).

Ocampo R. P. (2012, December). *History of Spacecraft Safety: U.S. Manned Space Program*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo R. P. (2013, March). *History of Spacecraft Safety: Soviet/Russian Manned Space Program*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo, R. P., & Klaus, D. M. (2013). A Review of Spacecraft Safety: From Vostok to the International Space Station. *New Space*, *1*(2), 73-80.

CHAPTER 2

PROBLEM STATEMENT

"…every manned spacecraft that leaves the earth . . . shall represent the best that dedicated and inspired men can create. We cannot ask for more; we dare not settle for less."

- Bob Gilruth, NASA Manned Spacecraft Center Director

2.1 History

Space flight is inherently risky. But how risky is too risky? Or conversely, how safe is safe enough? Questions like these have surrounded the space program since its inception. As early as Project Mercury, space officials were asking "how simple is safe?" (Swenson et al., 1966), "what does it mean to be reliable?" (Hacker & Grimwood, 1977), and "what…does good mean?" (Murray and Cox, 1989). In the years preceding the Space Shuttle Program, the question became: "[are] the risk[s] commensurate with the benefits?" (ASAP, 1978). Eventually, this question evolved into its present (and perhaps, most familiar) refrain: "how safe is safe enough?" (ASAP, 2009).

In the last decade, this question has grown in urgency. The Aerospace Safety Advisory Panel (ASAP)—the independent advisory panel tasked by Congress with evaluating NASA's safety performance—has posed the question "how safe is safe enough?" *in six of their last eight annual reports* (ASAP, 2009; ASAP, 2010; ASAP, 2011; ASAP, 2012; ASAP, 2014; ASAP, 2015).

This question remains both *critical* and *relevant* to human space flight because it must be answered constantly and consistently, on both a program-by-program and flight-by-flight basis; there is no comprehensive "right" answer. Different programs, flying different missions of various durations, may be willing to accept more or less risk (and more or less uncertainty within the assessment of risk). The Atlas Launch Vehicle (LV) was test flown 73 times before it was considered "safe enough" for crewed use; the Saturn V was test flown just *twice* before it sent humans to the moon (**Figure 6**).

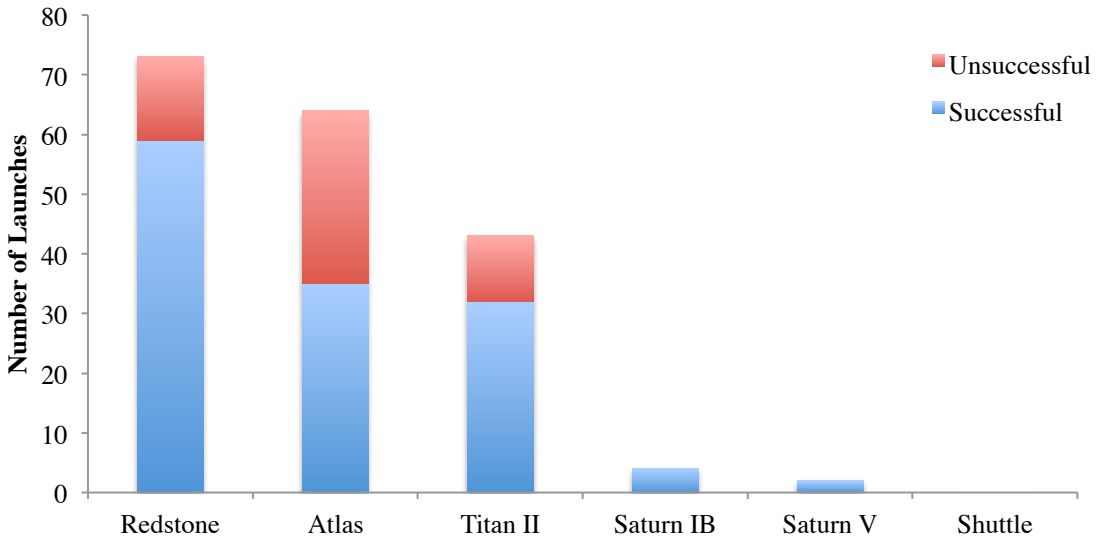


Figure 6: Number of successful and unsuccessful launches prior to first crewed flight for the various U.S. human space programs.

2.2 Sub-Questions

The *process* required to determine "safe enough," however, is relatively consistent, as it is the same process required to determine enough of *anything* (Ansoff, 1968). This process can be logically broken down into three steps: 1) define terminology, 2) characterize risk, and 3) determine "safe enough." Each of these three steps serves to

answer one or more related sub-questions, as depicted below (and described in detail in the paragraph that follows):

Step 1: Define Terminology

1a) What does it mean to be "safe?"

1b) What does it mean to be "unsafe?"

1c) What does "risk" mean?

Step 2: Characterize Risk

2a) How should risk be measured?

2b) How risky is space flight?

Step 3: Determine "Safe Enough"

3a) What does it mean to be "safe enough?"

3b) What is the minimum level of risk that can be achieved?

During step 1, relevant terminology is defined. Sub-questions such as what does it mean to be "safe?" (sub-question 1a), what does it mean to be "unsafe?" (sub-question 1b), and what does "risk" mean? (sub-question 1c) are presented and answered. Next, a metric for measuring risk is established. This metric serves to define the method of measurement (sub-question 2a), as well as the reference units for "safe" (e.g., number of successful launches, levels of failure tolerance, performance in a flight readiness review). Lastly, a criterion value (either quantitative or dichotomous) is assigned to "safe enough": for example, ***18*** successful launches, ***3*** levels of failure tolerance, ***successful completion*** of a flight readiness review. (sub-question 3b). This value is assigned after a review of the historic record to ensure it can be realistically achieved (sub-question 2b).

Once this value has been established, "safe enough" can be determined. If a spacecraft meets (or in certain cases, exceeds) the criterion, it is "safe enough"; if it does not meet the criterion, it is "not safe enough" (sub-question 3a).

SIDEBAR: The Luggage Analogy

The process of determining "safe enough" is perhaps better understood by analogy. Consider an airline that is looking to limit the amount of luggage their passengers bring on board an airplane. Before the airline can determine "how much luggage is too much luggage," they must first define what "luggage" is (step 1, sub-question a-c). This is not as trivial as it first may seem: a suitcase is an obvious candidate for luggage, but what about an article of clothing or bottle of water?

Next, a metric for measuring "luggage" must be established (step 2, sub-question a). Is it by weight in pounds? By cubic volume? By number of items? If the airline decides to measure luggage by weight in pounds, the units associated with "luggage" and "too much luggage" must also be measured by weight in pounds. However, the actual *value* associated with "too much luggage" must still be defined.

Theoretically, this "too much luggage" value could be selected at random. An astute airline, however, will first review the *actual* amount of luggage (in pounds) their passengers have carried in the past, as well as the amount of weight (again, in pounds) their aircraft are expected to carry in the future (step 2, sub-question b). In this manner, the value that is ultimately selected for "too much luggage" can more accurately reflect reality (step 3, sub-question a-b).

2.3 Complications

Although these sub-questions may seem straightforward in theory, answering them can be difficult and contentious in practice, as the inherent uncertainties associated with measuring "safe" are complicated by the subjective challenges of determining "enough." These complications—*uncertainty in terminology*, *subjectivity in the choice of metrics*, *uncertainty in the validity of the metrics*, *uncertainty in the measurement itself*, and *subjectivity in the acceptance of the measurement*—must be avoided, eliminated, or accounted for when attempting to objectively evaluate "safe enough." These complications are described in detail below.

2.3.1 Complication 1: Uncertainty in Terminology

Although the definitions of "safe," "unsafe," and "risk are *generally* well agreed upon within the English language, there are certain instances where the first two terms—"safe" and "unsafe" overlap, particularly when they are applied retroactively. For example, in the years prior to the *Challenger* accident, the Space Shuttle was described as becoming increasingly "***unsafe***" (U.S. House of Representatives, 1986). However, each of the launches prior to *Challenger* also resulted in the crew's "***safe"*** return. This contradiction in terms suggests the words "safe" and "unsafe," as currently defined, are vague and potentially misleading. This complication directly affects sub-questions 1a, *what does it mean to be "safe?"*; 1b, *what does it mean to be "unsafe?";* and 1c, *what does "risk" mean?*

2.3.2 Complication 2: Subjectivity in the Choice of Metrics

Unlike other physical variables, such as mass or length, "safe" cannot be measured empirically; it must be abstracted from the spacecraft and its interaction with the environment (ASAP, 2002). How this abstraction should proceed is a subjective choice: should it be based on the rate of successful launches? Or should it be quantified using probabilistic calculations? Or should some other metric be applied?

The process of choosing a "safe" metric can be highly contentious: Apollo engineers were "deep[ly] and irreconciab[ly]" divided as to whether "safe" should be estimated using statistical or actuarial analysis (Murray and Cox, 1989). Later Space Shuttle engineers faced a similar struggle when deciding between Failure Modes and Effects Analysis (FMEA) or Probabilistic Risk Analysis (PRA) as the primary means of measuring "safe" post-*Challenger* (Feynman, 1986; Fragola, 1996; Slay, 1988; Vaughan, 1996). Such a complication directly affects the answer to sub-question 2a, *how should risk be measured?*

2.3.3 Complication 3: Uncertainty in the Validity of the Metrics

The choice of a metric is further complicated by the fact that no metric can ever serve as a perfect (e.g., exact) proxy for "safe." For example, counting the number of redundant components within a system is a reasonable approach to quantifying "safe," one loosely (and more qualitatively) employed by the Space Shuttle Program in its early years under the auspices of the Critical Items List (CIL) (Slay, 1988). However, the addition of redundant components can sometimes add complex and unpredictable failure modes to a system—which in turn can lead to an overall reduction in "safe" (Ocampo,

2014). In these limited instances, counting redundant components actually serve as a specious indicator of "safe," thereby invalidating the metric.

Even relatively simple metrics, such as dividing the number of successful launches by the number of total launches, are not immune to this complication. Consider a spacecraft that has been successfully launched one time (and one time only). This spacecraft would have a mathematically perfect safety record, but could not in good conscience be described as perfectly "safe." This complication affects how sub-question 2a, *how should risk be measured?* is answered.

2.3.4 Complication 4: Uncertainty in the Measurement Itself

Few metrics (if any) can quantify "safe" with perfect precision; even something as simple as counting redundant parts can generate uncertainty. The O-rings on the Space Shuttle Solid Rocket Boosters (SRBS), for example, were originally classified as "criticality 1R"—meaning they were considered redundant to catastrophic failure. However, this classification was later changed to "criticality 1" (e.g., *not* redundant) in 1982, when engineers realized that leakage of the primary O-ring during certain limited phases of flight was actually a single-point failure (Rogers et al., 1986). Notably, this classification change occurred in the absence of any modifications to the design, suggesting that even a simple metric for measuring "safe"—like redundancy counts—can have uncertainty associated with its measurement.

The existence of such uncertainty must be accounted for when attempting to measure "safe" or determine "safe enough." Probabilistic Risk Assessment achieves this

by generating *two* measurements: a mean estimate of risk *and* an estimate of risk uncertainty. This uncertainty affects sub-question 2b, *how safe is space flight?*

2.3.5 Complication 5: Subjectivity in the Acceptance of the Measurement

Determining whether a spacecraft's measured level of "safe" is "safe *enough*" is ultimately a subjective decision. This does not mean, however, that it is necessarily an easy or inconsequential one. In order to accept a spacecraft as "safe enough," anticipated mission benefits must be shown to demonstrably outweigh the potential for mishap (ASAP, 2014).

This subjectivity readily influences whether a spacecraft is deemed "safe enough," thereby affecting sub-question 3a, *what does it mean to be safe enough?*, and the answer to sub-question 3b, *what is the minimum level of risk that can be achieved?*

2.4 Objectives

The process of answering each sub-question, while navigating their associated complication(s), is the primary goal of this thesis, as this serves to help answer the primary question, "how safe is safe enough?" The details of this process are described briefly below, addressed succinctly in **Table 3**, and expanded upon in Chapters 3, 4, and 5.

Chapter 3—Definitions and Framework: This chapter defines three key elements: "safe," "unsafe," and "risk." These definitions are derived from (and consistent with) current NASA terminology, real-world examples, and empirical practice, and serve to

provide a framework (defined here as a "risk spectrum") for characterizing and predicting risk in Chapter 4 and Chapter 5. This chapter addresses sub-questions 1a, *what does it mean to be "safe?"*; 1b, *what does it mean to be "unsafe?";* and 1c, *what does "risk" mean?*

Chapter 4—Characterizing Space Flight Risk: This chapter establishes a metric for measuring risk" based on (and consistent with) the definitions established in Chapter 3. Historic safety records of crewed spacecraft (Space Shuttle, Soyuz) are then measured against this metric, and presented in the context of the "risk spectrum" from an *absolute* perspective. These safety records are then compared with the safety records of various modes of transportation (automotive, rail, boating general aviation, commercial aviation) and adventure sport activities (skydiving, mountaineering, SCUBA diving) in order to characterize the *relative* risk of space flight. This chapter serves to answer sub-questions 2a, *how should risk be measured*?; and 2b, *how risky is space flight?*

Chapter 5—Determining Safe Enough: This thesis closes by establishing a definition for "safe enough" which is consistent with the terminology and framework defined in Chapter 3 and complaint with the risk metric established in Chapter 4. A novel means of predicting risk, built upon risk and usage data collected in Chapter 4, is then established. This risk heuristic is independent of any specific design or program, and as such, can be used to predict achievable levels of risk in the absence of detailed spacecraft data. Based on these predictions, anticipated spacecraft risk can either be accepted as "Safe Enough,"

or rejected as "Not Safe Enough." This chapter serves to answer sub-questions 3a, *what does it mean to be "safe enough"?* and 3b, *what level of risk is achievable?*

Step:	1: Define Terminology			2: Characterize Risk		3: Determine "Safe Enough"	
Sub-question:	1a) What does it mean to be "safe?"	1b) What does it mean to be "unsafe?"	1c) What does "risk" mean?	2a) How should risk be measured?	2b) How risky is space flight?	3a) What does it mean to be "safe enough"?	3b) What is the minimum level of risk that can be achieved?
Complication:	*Subjectivity in defining terminology*			*Subjectivity in the choice of metrics*, *Uncertainty in the validity of the metrics*	*Uncertainty in the measurement itself*	*Subjectivity in the acceptance of the measurement*	
Addressed in:	Chapter 3—Definitions and Framework			Chapter 4—Characterizing Space Flight Risk		Chapter 5—Determining Safe Enough	

Table 3: The process of determining "safe enough" can be broken down into three steps, each with their own set of sub-questions and complications. Each of the sub-questions is addressed in the chapters that follow.

2.5 Syntax

In this thesis, when a quoted word is capitalized (e.g. "Safe", "Unsafe", "Safe Enough", Not Safe Enough") that word is referencing a specific system state. When it is not capitalized (but still quoted), the word is referencing its definition.

2.6 Immediate Forward Work

No spacecraft will ever be perfectly safe. As such, the intent of this thesis is to define frameworks, characterize risk, and help establish thresholds that can be used to determine whether a spacecraft is "safe enough."

2.7 Related Publications, Presentations, and Posters

Ocampo R. P. (2011). *Human Rating Space Systems: How Safe is Safe Enough?* Presented at ASGSB/AIAA Student Panel, Washington D.C.

Klaus, D., Fanchiang, C., & Ocampo, R. (2012). Perspectives on Spacecraft Human-Rating. In *42nd International Conference on Environmental Systems* (p. 3419).

Ocampo, R. P., & Klaus, D. M. (2013). A Review of Spacecraft Safety: From Vostok to the International Space Station. *New Space*, *1*(2), 73-80.

Klaus, D., Ocampo, R., & Fanchiang, C. (2014, March). Spacecraft human-rating: Historical overview and implementation considerations. In *Aerospace Conference, 2014 IEEE* (pp. 1-7). IEEE.

Ocampo, R. P. (2014). Limitations of Spacecraft Redundancy: A Case Study Analysis. 44th International Conference on Environmental Systems.

Ocampo R. P. (2014, July). *The Limitations of Spacecraft Redundancy*. Presented at 44th International Conference on Environmental Systems (ICES), Tucson AZ.

Ocampo, R. P., & Klaus, D. M. (2015, October). *Human Space Flight Safety*. Poster session presented at the FAA COE CST 5th Annual Technical Meeting, Washington DC.

CHAPTER 3

DEFINITIONS AND FRAMEWORK

"When we first started [flying in space], people would say things like, well the spacecraft's got to be 'good'. But what the hell does 'good mean?"

- Glynn Lunney, Gemini and Apollo Flight Director

3.1 Objective

Before "safe enough" can be determined, "safe" must first be defined. As such, this chapter has two major objectives:

1. Develop a definition of "safe" that is consistent with NASA terminology, real-world examples, and empirical practice. In the process of defining "safe," its antithesis, "unsafe" and its descriptor, "risk" are also defined.
2. Create a conceptual framework, known as the "risk spectrum," which can be used to characterize risk in Chapter 4 and help to determine "safe enough" in Chapter 5.

3.2 Background

Although the colloquial definition of "safe" is generally well agreed upon, there is evidence to suggest that the term, when used in the context of engineering safety analyses, may actually be quite equivocal. In their 1978 annual report, the Aerospace Safety Advisory Panel (ASAP) stated that one of the primary obstacles to determining

"safe enough" stems from the ambiguous use of the term "safe" in the English lexicon. They wrote:

> *The very nature of safety determinations and the wide-spread confusion about the nature of safety decisions would be dispelled if the very meaning of the term were clarified* (ASAP, 1979).

This "need for clarification" stems from the fact that both "Safe" and "Unsafe"—two terms with seemingly antithetical definitions—are often readily ascribed to the same spacecraft. The U.S. House of Representatives, in their investigation of the Space Shuttle *Challenger* accident, wrote that over the course of the first 24 launches (e.g., all launches prior to *Challenger*), the Space Shuttle was becoming "increasingly *unsafe*" [emphasis added] (U.S. House of Representatives, 1986). However, each of these 24 launches resulted in the crew's *safe* return. Moreover, the mission that directly preceded *Challenger* successfully launched a sitting politician, Representative Bill Nelson.

In a similar vein, NASA made the decision to retire the Space Shuttle after the *Columbia* disaster in part because the system's age suggested the vehicle was growing increasingly unsafe (Day, 2011). Nevertheless, the shuttle flew 22 times after *Columbia*, with each mission resulting in the crew's safe return; news reports of the shuttle's final flight even described *Atlantis* as returning "her crew home *safely*" [emphasis added] (Bergin, 2011).

These examples are not intended to construe the Space Shuttle as having been "Safe" or "Unsafe" but rather to exemplify the overlapping (and therefore, sometimes equivocal) use of the two terms within the English language. Therefore, before "safe

enough" can be determined, unambiguous definitions of "Safe" and "Unsafe" must first be established.

3.3 Definitions

3.3.1 Baseline Definitions

This thesis utilizes NASA's current definition of safety as an initial baseline because it provides a clear and unequivocal description of what is safe and what is not. NASA defines safety as:

> *Freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment* (NASA, 2008b).

Given this definition, a spacecraft can exist in only one of two exclusive states, "Safe" and "Unsafe":

Safe (NASA Baseline Definition): Spacecraft is free from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment.

Unsafe (NASA Baseline Definition): Spacecraft is **NOT** free from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment.

Since no spacecraft can ever truly be free from "conditions…" that can cause harm (Slay, 1988; ASAP, 1981), no spacecraft can ever be considered "Safe"[2]. Therefore, all systems are *by NASA's definition* "Unsafe." Moreover, all spacecraft are *uniformly* "Unsafe" by this classification system, as the discrete definition provided by NASA does not differentiate between varying *degrees* of "Unsafe."

3.3.2 Evolving Definitions

Such prescribed uniformity, however, is contraindicated by a number of real-world examples. Consider:

- A spacecraft that exposes its crew to *10* catastrophic conditions(e.g., conditions that can result in "fatal injury, loss of vehicle, or permanently disabling injury [NASA, 2008b]) is understood to be "more unsafe" than a spacecraft that exposes its crew to *1* catastrophic condition (assuming each condition is equally likely to occur).

- A spacecraft that exposes its crew to 1 *likely* catastrophic condition (e.g., 99% likelihood) is generally considered to be "more unsafe" than a spacecraft that exposes its crew to 2 *unlikely* catastrophic conditions (e.g., each condition has a 1% likelihood).

---

[2] This is not to say that such "conditions" will *always* cause harm, but rather that "conditions" will always exist in real-world spacecraft with the *potential to cause harm*.

Therefore, the definitions of "Safe and "Unsafe" must be amended to properly account for both the number and likelihood of the "conditions" facing the crew:

Safe (Evolved Definition 1): Spacecraft is free from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment. **Given that no practical (e.g., non-theoretical) system can ever be free of such "conditions," this state is unachievable.**

Unsafe (Evolved Definition 1): **One or more conditions can occur** that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment. **The likelihood of any one of these conditions occurring is directly proportional to the degree to which the system is "Unsafe."**

The severity of the "conditions…" must also be specified within the definitions of "Safe" and "Unsafe" to ensure each spacecraft is assessed against an equivalent standard of comparison. A spacecraft that exposes its crew to 1 *catastrophic* condition is implicitly understood to be "more unsafe" than a spacecraft that exposes its crew to 1 *critical* condition (e.g., one that can result in "severe injury or occupational illness, or major property damage to facilities, systems, or flight hardware" [NASA, 2008b]))—assuming both conditions are equally likely. This thesis focuses specifically on *catastrophic* conditions, as these conditions tend to be of primary concern to NASA and

other space agencies[3]; as such, the definition of "Safe" and "Unsafe" are further modified as follows for this framework:

Safe (Evolved Definition 2): Spacecraft is free from all **catastrophic** conditions. Given that no practical (e.g., non-theoretical) system can ever be free of such "conditions," this state is unachievable.

Unsafe (Evolved Definition 2): One or more **catastrophic** conditions can occur. The likelihood of any one of these catastrophic conditions occurring is directly proportional to the degree to which the system is "Unsafe."

3.3.3 Final Definitions

The terminology used to define "Safe" and "Unsafe" above is precise but unwieldy. To simplify these definitions, the term "hazard" will be used instead of "conditions…," as the two terms are virtually synonymous per NASA's definition:

> A hazard is "*a state or a set of conditions, internal or external to a system that has the potential to cause harm*" (NASA, 2008b).

---

[3] It should be noted that alternative definitions of safety can employ different levels of severity (e.g., critical, severe, moderate, or minor) without compromising the general concept of "Safe" and "Unsafe" described herein. The key to the definition's utility within a "Safe Enough" framework is that the severity level is specified and preserved throughout the analysis. This helps to ensure spacecraft are assessed against an equivalent standard of comparison.

Additionally, because NASA defines "risk" as:

> *the combination of the probability (qualitative or quantitative) of experiencing an undesirable event [e.g., hazard], and the uncertainties associated with the probabilities and consequences* (NASA, 2008b),

"degrees of unsafe" can (and will) be articulated as "risk" throughout this thesis. A final definition of "Safe," "Unsafe" and "Risk" is found below; for a more detailed description of the evolution of the terms, see **Table 4**.

Safe (Evolved Definition 3—Final Definition): Spacecraft is free from all catastrophic hazards. Given that no practical (e.g., non-theoretical) system can ever be free of such hazards, this state is unachievable.

Unsafe (Evolved Definition 3—Final Definition): One or more catastrophic hazard(s) can occur. The likelihood of any one of these catastrophic hazard(s) occurring is directly proportional to the degree to which the system is "Unsafe."

Risk (Final Definition): The degree to which a system is "Unsafe."

Version	Reason for Update	Definition of Safe	Definition of Unsafe
Baseline (via NASA NPR 8715.3C)	N/A	Spacecraft is free from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment.	Spacecraft is **NOT** free from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment.
NASA Baseline Def.	Since no spacecraft can ever be free from conditions that can cause harm, no spacecraft can ever be considered "Safe."	Spacecraft is free from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment. *Given that no practical spacecraft can ever be free of such "conditions," this state is unachievable.*	NO CHANGE
Evolved Def. 1	Since no spacecraft can ever be free from conditions that can cause harm, all spacecraft must be considered "Unsafe". However, not all systems are uniformly "Unsafe". Rather, there are varying degrees of unsafe, affected by the number and likelihood of the conditions.	NO CHANGE	*One or more conditions can occur that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment. The likelihood of any one of these conditions occurring is directly proportional to the degree to which the system is "unsafe."*
Evolved Def. 2	The severity of the conditions must be specified to ensure that spacecraft are assessed against equivalent standards. The definition listed in this thesis specifies "catastrophic" conditions, as these are of primary concern to space agencies.	Spacecraft is free of all *catastrophic* conditions. Given that no practical system can ever be free of such "conditions", this state is unachievable.	One or more *catastrophic* conditions can occur. The likelihood of any one of these *catastrophic* conditions occurring is directly proportional to the degree to which the system is "Unsafe."
Evolved Def. 3 (Final Def.)	NASA defines a hazard as a "state or a set of conditions, internal or external to a system that has the potential to cause harm". Therefore, hazard can replace "those conditions…," thereby simplifying the definitions.	Spacecraft is free of all catastrophic *hazards*. Given that no practical system can ever be free of such hazards, this state is unachievable.	One or more catastrophic *hazards* can occur. The likelihood of any one of these catastrophic *hazards* occurring is directly proportional to the degree to which the system is "Unsafe".
NOTE: NASA defines risk as "the combination of the probability (qualitative or quantitative) of experiencing an undesirable event, and the uncertainties associated with the probabilities and consequences." (NASA 2008b) Therefore, "degrees of unsafe" will be articulated as "degrees of risk" throughout the remainder of this thesis; however, the final definition of "Unsafe" does not change.			

Table 4: Evolution of "Safe" and "Unsafe" definitions. When new aspects of the definitions are added, they are depicted in blue italics.

3.4 Conceptual Framework

The likelihood that a spacecraft will experience one or more catastrophic hazards naturally ranges from 0% to 100%. Therefore, risk (previously, "degrees of Unsafe") must *also* exist as a spectrum of values ranging from 0% to 100% (see **Figure 7**). The lower limit of this spectrum, 0%, represents the "safe" state; as such it is not practically achievable. The rest of the spectrum (risk > 0%) represents the "unsafe" state. Identifying a spacecraft's exact position within this framework (e.g., measuring its risk) is described in detail in Chapter 4.

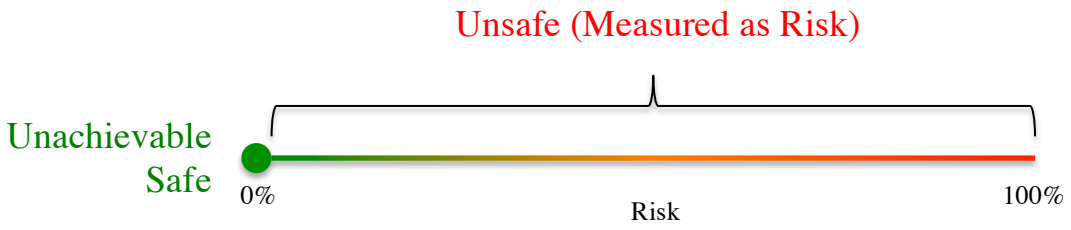


Figure 7: Risk (e.g., the degree to which a spacecraft is "unsafe") can be characterized as a spectrum per the figure above. A spacecraft on the right side of the spectrum has relatively high risk, while a spacecraft on the left side of the spectrum has relatively low risk. In this spectrum, a risk value of 0% is equivalent to the "Safe" state, which is considered unachievable by this framework.

Although a similar framework for defining risk has been alluded to elsewhere (Dezfuli, 2010; Stamatelatos, 2010), to the author's knowledge, the framework presented here is the first to be derived from the bottom up, using first-order logic, based strictly on the essential components of the question, "how safe is safe enough?"

3.5 Chapter Summary

This chapter unequivocally defined three key terms: "safe," "unsafe" and "risk." These definitions were derived from NASA terminology, real-world examples, and empirical practice; as such, uncertainty in the terminology (complication 1) was mitigated to the maximum extent practicable. These terms served to establish a conceptual "risk spectrum" framework, which can be used to characterize and compare spacecraft risk.

In the process of defining terminology and establishing the risk spectrum framework, three sub-questions were addressed: sub-question 1a, *what does it mean to be "safe?"*; sub-question 1b, *what does it mean to be "unsafe?";* and sub-question 1c, *what does "risk" mean?*

3.6 Immediate Forward Work

Given the terms and framework defined here, a metric for measuring risk can now be established (Chapter 4). Using this metric, past spacecraft risk can be characterized from both an absolute and relative perspective (Chapter 4).

3.7 Related Publications, Presentations, and Posters

Ocampo, R. P., & Klaus, D. M. (2012, July). *Defining a Safety Index for Human Spacecraft Design*. Poster session presented at the AIAA 42nd International Conference on Environmental Systems (ICES), San Diego, CA.

Ocampo R. P. (2013, February). *Evaluating Safety from a Quantitative Perspective*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo R. P. (2014, January). *Risk and Safety: Refining NASA's Nomenclature*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo, R. P., & Klaus, D. M. (2015, October). *Human Space Flight Safety*. Poster session presented at the FAA COE CST 5th Annual Technical Meeting, Washington, D.C.

Ocampo, R. P., & Klaus, D. M. (2016). A Quantitative Framework for Defining “How Safe is Safe Enough” in Crewed Spacecraft. *New Space*. *Manuscript in press*.

CHAPTER 4

CHARACTERIZING SPACE FLIGHT RISK

"The joke that made the rounds of NASA was that the Saturn V had a reliability rating of .9999. In the story, a group from headquarters goes down to Marshall and asks Wernher von Braun how reliable the Saturn is going to be. Von Braun turns to four of his lieutenants and asks, "Is there any reason why it won't work?" to which they answer: "Nein." "Nein." "Nein." "Nein." Von Braun then says to the men from headquarters, "Gentlemen, I have a reliability of four nines."

- Apollo: The Race to the Moon (Murray & Cox, 1989, p. 87)

4.1 Objective

In order to effectively determine "safe enough," a detailed understanding of space flight risk is required. Consequently, this chapter has two major objectives:

1. Establish a metric for measuring risk that is consistent with the terminology and framework established in Chapter 3.
2. Using this metric, characterize past spacecraft risk from both an absolute and relative perspective.

4.2 Risk Metrics

In order to characterize the degree to which a spacecraft is "Unsafe," a metric for measuring risk must first be developed or established. Given the framework presented in Chapter 3 (and depicted in Figure 8), this metric must be capable of calculating the likelihood that a given spacecraft will experience one or more catastrophic hazards. Only two metrics to date can effectively perform this calculation: actuarial analysis and Probabilistic Risk Analysis (PRA)[4]. These metrics are described in detail below.

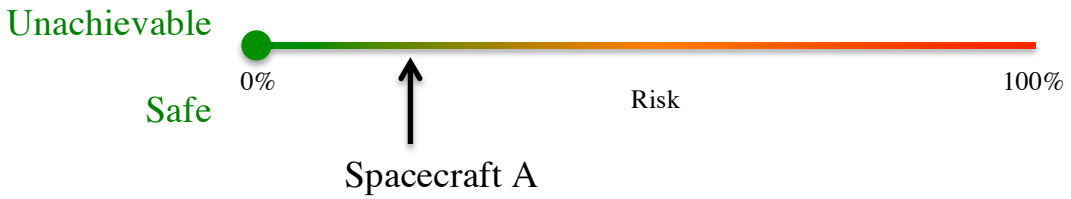


Figure 8: In order to determine a spacecraft's position on the risk spectrum, a method for measuring risk is required. Both probabilistic and actuarial analysis can be used to perform this measurement.

4.2.1 Actuarial Analysis

Actuarial analysis measures risk by summing the total number of incidents (e.g., fatal accidents, fatalities, etc.) for a given exposure period or characteristic (trips, passengers, miles, hours, etc.). An actuarial analysis, for example, would describe the Space Shuttle as having had

---

[4] NASA relies on several different metrics (in addition to actuarial analysis and PRA) to evaluate risk. However, these methods, which include Hazard Analysis (HA), Fault Tree Analysis (FTA), and Failure Modes and Effects Analysis (FMEA) (see Appendix B), can only evaluate risk at a *qualitative* level; they are therefore incompatible with the quantitative risk spectrum presented here.

2 fatal accidents in 135 flights (or equivalently, 1 fatal accident in 67.5 flights). Because actuarial analysis is a count of historic data, there are no uncertainty values associated with its measurements.

4.2.2 Probabilistic Risk Analysis

Conversely, PRA estimates risk by developing a mathematical-logical model of a physical system (Bogumil, 1982). Events within the model that can lead to a catastrophic accident are assigned probabilities; these probabilities are then collated to identify an overall probability of Loss Of Crew [p(LOC)]. This type of analysis generates a mean estimate of risk, as well as an estimate of uncertainty. Towards the end of its operational lifetime, the Space Shuttle mean p(LOC) was calculated to be 1 in 90, with an error factor of 1.4 (Hamlin et al., 2011).

4.2.3 Down-Selecting a Metric

Probabilistic metrics tend to be better predictors of risk during the early stages of program development (or after any substantial design changes), when there are few, if any, relevant launches to assess actuarially. Given that human space flight is still a relatively nascent industry (despite its 50+ year history), most human space flight programs tend to rely on PRA, rather than actuarial analysis, to assess spacecraft risk.

The benefit of actuarial analysis is that it—unlike PRA—has zero uncertainty associated with its measurement values. Moreover, actuarial analysis is a highly valid means of assessing risk *once sufficient data points have been collected*. Consider a two-sided coin: it may be impossible to determine if the coin is balanced after a single flip, but after 20 flips, balance can

be reasonably inferred (assuming the flips result in a roughly equal number of heads and tails—see Figure 9).

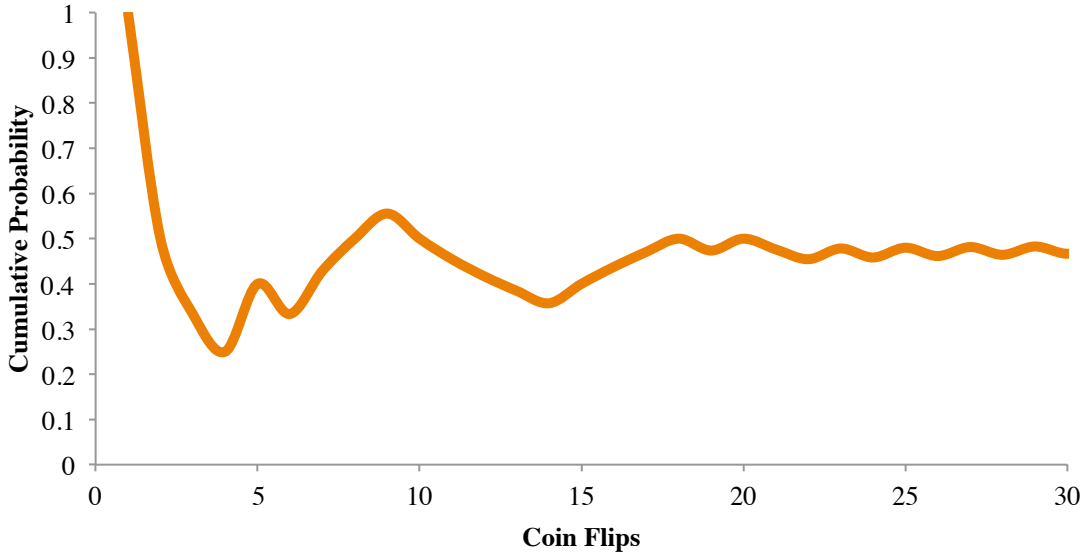


Figure 9: Cumulative probability of randomly flipping heads for a balanced coin. During the first several flips, it is difficult to determine whether the coin is actually balanced (e.g., cumulative probability equal to 50%). However, as the number of flips increases, balance can be more readily inferred.

PRA is also a relatively new and generally proprietary metric for assessing risk within the aerospace industry (Fragola, 1996). As such, there are very few probabilistic data points available to the academic community that can be used to represent space flight risk.

Given these circumstances, space flight risk will be characterized using actuarial analysis in this thesis. This decision (which relates to complication 2, *subjectivity in the choice of*

metrics) is also justified by the fact that the metric serves to resolve complication 4, *uncertainty in the measurement itself* by producing values with zero uncertainty[5].

4.3 Absolute and Relative Risk

The remainder of this chapter is devoted to characterizing the safety records (i.e., the actuarial risk) of the Space Shuttle and Soyuz spacecraft. These two spacecraft account for 88% of all orbital human space flights to date; therefore, their safety records are likely to be representative of current human space flight risk.

Space Shuttle and Soyuz safety records are then compared to the safety records of various modes of transportation (automotive, rail, boating general aviation, commercial aviation) and adventure sport activities (skydiving, mountaineering, SCUBA diving). These *relative* risk comparisons, though unnecessary for determining *absolute* spacecraft risk, are used to establish a method for determining "safe enough" in Chapter 5[6].

4.4 Reference Units

The safety records of each of the different activities are presented using six different units of measurement, namely: fatal accidents per vehicle-trip, fatal accidents per vehicle-hour, fatal accidents per vehicle-mile, fatalities per person-trip, fatalities per person-hour, and fatalities per

---

[5] A means of overcoming complication 3, *uncertainty in the validity of the metric*—which in the case of actuarial analysis, stems from space flight's low launch rates—is discussed in detail in Chapter 5.

[6] Additionally, there is strong evidence to suggest relative risk comparisons may be a more effective means of conveying risk to uninformed individuals (Ocampo & Klaus, *manuscript in prep*).

person-mile[7]. These reference units were chosen to represent risk because they are units most commonly used by NASA and the transportation industry at large.

These six reference units can be categorized as either *vehicle-centric* or *person-centric*:

Vehicle-centric: Vehicle-centric reference units (*fatal accidents per vehicle-trip*, *fatal accidents per vehicle-hour*, *fatal accidents per vehicle-mile*) emphasize the risk associated with a given *vehicle*. Vehicle-centric reference units are not affected by the number of participants on board the vehicle, and therefore do not distinguish between events which kill a fraction of the participants and events which kill all on board. Historically, NASA has presented risk (both actuarial and probabilistic) using vehicle-centric reference units (e.g., Loss of Crew). Given the focus of vehicle-centric reference units, they are not amenable to presenting the safety records of adventure sport activities, which rarely involve the use of distinct vehicles.

Person-centric: Person-centric reference units (*fatalities per person-trip*, *fatalities per person-hour*, *fatalities per person-mile*) emphasize the *individual* risk associated with a given activity, specifically accounting for the fact that risk may vary from individual to individual within the same vehicle.

---

[7] In the luggage analogy presented in the introduction, these reference units are analogous to the different units of weight in which luggage can be measured, e.g., pounds or kilograms. For a complete description of these terms, see Appendix C.

4.5 Methodology

Fatality and exposure data for 8 of the activities reviewed here—mountaineering on Denali (Alaska), mountaineering on Mt. Rainier (Washington), driving a personal automobile on U.S. roads, travel aboard Amtrak passenger trains, boating within U.S. waters, flights aboard U.S. part 91 (general) aviation, flights aboard U.S. part 121 (scheduled airline) aviation, and flights aboard the Space Shuttle—were aggregated from several different U.S. government sources, including the National Park Service (NPS), the National Highway Traffic Safety Administration (NHTSA), the Federal Highway Administration (FHWA), the Federal Railroad Administration (FRA), the U.S. Coast Guard (USCG), the Federal Aviation Administration (FAA), the Bureau of Transportation Statistics (BTS), and the National Aeronautics and Space Administration (NASA). The anticipated risk associated with NASA's Commercial Crew Program (CCP)—derived specifically from NASA requirements—was also included in the analysis for context, even though the value is probabilistic as opposed to actuarial.

For those 4 activities that were not directly regulated by the U.S. government during the time periods reviewed here (e.g., climbing Mt. Everest, skydiving, SCUBA diving, and flights aboard Soyuz), statistics were gathered either from the activity's governing body or from other non-U.S. government agencies. Concerted efforts were made to ensure fatality and exposure data sources were consistent within each activity so as to best maintain the precision of each calculated reference unit.

Wherever possible, raw fatality and exposure data for each activity were aggregated over a 5-year time period to help minimize the effects of outlier years. With very few exceptions, these time periods represent the most recent years for which data was available so as to help

ensure the *currency* of risk comparisons. The specific time periods that were assessed for this thesis are listed in Table 5**.**

This raw fatality and exposure data was then processed to generate rates of fatal accidents per vehicle-trip, fatal accidents per vehicle-hour, fatal accidents per vehicle-mile, fatalities per person-trip, fatalities per person-hour, and fatalities per person-mile.

In order to verify whether risk differences between activities were statistically significant, a Chi-squared test was performed within each reference unit. If the riskiest activity was found to be significantly greater than the second riskiest activity, it was inferred to be significantly greater than *all* of the activities.

4.6 Results

Raw fatality and exposure characteristics for each of the 12 different activities are listed in Table 5. Processed data is depicted in Figure 10 **-** Figure 15 and described in sections 4.6.1 – 4.6.6.

Activity	Fatalities	Fatal Accidents	Vehicle-Trips	Vehicle-Hours	Vehicle-Miles	Person-Trips	Person-Hours	Person-Miles
U.S./Canada SCUBA (Insured Divers) *2000-2006*	187	187	N/A	N/A	N/A	2.67×10^{7}	N/A	N/A
Skydiving *2010-2014*	113	113	N/A	N/A	N/A	1.56×10^{7}	N/A	N/A
Mt. Everest *2005-2009*	23	23	N/A	N/A	N/A	2.28×10^{3}	N/A	N/A
Denali *2010-2014*	16	16	N/A	N/A	N/A	6.03×10^{3}	N/A	N/A
Mt. Rainier *2006-2010*	3	3	N/A	N/A	N/A	4.96×10^{4}	N/A	N/A
Automobile *2009*	33,883	30,862	2.34×10^{11}	*7.48×10^{10}*	2.25×10^{12}	3.27×10^{11}	N/A	3.30×10^{12}
Amtrak Rail *2010-2014*	19	1	5.66×10^{5}	3.92×10^{6}	2.04×10^{8}	1.47×10^{8}	N/A	3.37×10^{10}
Boating *2012*	651	578	2.44×10^{8}	1.39×10^{9}	N/A	5.86×10^{8}	3.58×10^{9}	N/A
U.S. Part 91 Aviat. *2003-2007*	2,957	1,583	1.91×10^{8}	1.22×10^{8}	*1.22×10^{10}*	*3.82×10^{8}*	N/A	N/A
U.S. Part 121 Aviat. *2008-2012*	57	5	4.88×10^{7}	9.02×10^{7}	3.85×10^{10}	3.65×10^{9}	*5.71×10^{9}*	2.86×10^{12}
Soyuz *1967-March 2015*	4	2	125	3.44×10^{5}	5.92×10^{9}	3.18×10^{2}	9.51×10^{5}	1.64×10^{10}
Shuttle *1981-2011*	14	2	135	3.19×10^{4}	5.43×10^{8}	8.17×10^{2}	2.00×10^{5}	3.40×10^{9}
NASA Com. Crew *Anticipated*	N/A	1	200	N/A	N/A	N/A	N/A	N/A

Associated References	
SCUBA	Denoble et al., 2011
Skydiving	United States Parachute Association
Mt. Everest	Salisbury & Hawley, 2011
Denali, Mt. Rainier	National Park Service
Automobile	Bureau of Transportation Statistics, National Highway Traffic Safety Administration, Santos et al., 2011
Amtrak	Federal Railroad Administration, Wikipedia - *List of rail accidents (2010-present)*, Amtrak, National Association of Railroad Passengers
Boating	United States Coast Guard
Part 91 Aviation	Bureau of Transportation Statistics, Federal Aviation Administration
Part 121 Aviation	Bureau of Transportation Statistics, National Transportation Safety Board
Soyuz	Wikipedia- *List of Soviet manned space missions*, *List of Soviet manned space missions*
Shuttle	Wikipedia - *List of Space Shuttle missions*
Commercial Crew	National Aeronautics and Space Administration

Table 5: Fatality and exposure data for space flight, terrestrial transportation, and adventure sport activities. Exposure data that was estimated by the author is listed in blue italics and described below.

A complete set of exposure data could not be identified in the literature, as not all activities maintain records for each exposure type. Certain combinations of activities and exposure types are simply not amenable to record keeping, either because the measurements are difficult to collect (e.g., Mt. Everest person-miles) or are of little interest to the activity's governing body (e.g., skydiving person-hours). Other combinations of activity and exposure measurements are nonsensical; there are generally no vehicles involved in SCUBA diving, so there is no way to measure SCUBA diving vehicle-hours.

In some cases however, exposure data could be readily estimated from previously identified exposure characteristics. For example, vehicle-miles could be approximated from vehicle-hours if the average speed of the vehicle was known. For this thesis, the following 4 exposure characteristics were estimated:

Automobile Vehicle-Hours: Automobile vehicle-hours were estimated by dividing vehicle-miles by 30 mph—the estimated average speed of an automobile for all automobile trips. A sensitivity analysis of this estimate, using values ranging from 10 mph to 60 mph, did not affect the automobile's relative risk ranking in terms of fatal accidents per vehicle-hour.

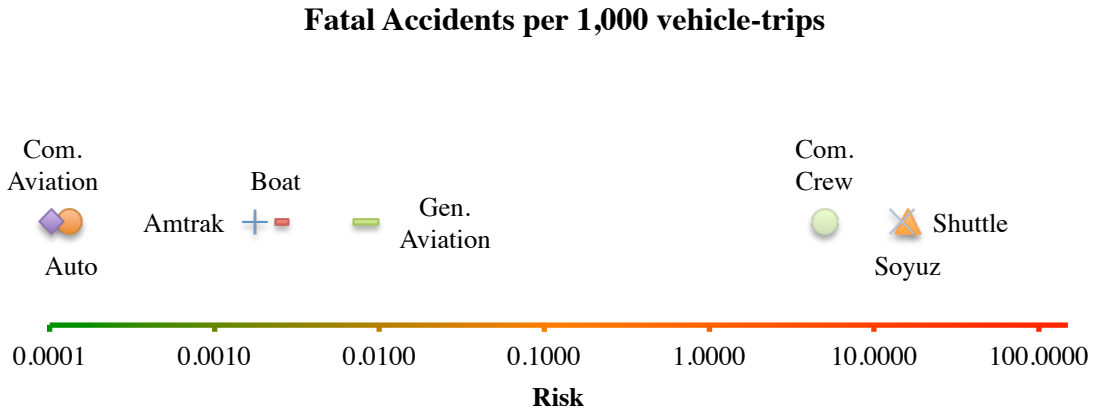
Part 91 (General) Aviation Vehicle-Miles: General Aviation (GA) vehicle-miles were estimated by multiplying vehicle-hours by 100 mph—the estimated average speed of a General Aviation aircraft. A sensitivity analysis of this estimate, using values ranging from 50 mph to 200 mph, did not affect general aviation's relative risk ranking in terms of fatal accidents per vehicle-mile.

Part 91 (General) Aviation Person-Trips: General Aviation person-trips were estimated by multiplying the total number of trips by 2—the estimated average number of people on board each GA aircraft during each trip. A sensitivity analysis of this estimate, using values ranging from 1 to 6 passengers, did not affect general aviation's relative risk ranking in terms of fatalities per person-trip.

Part 121 (Scheduled Airline) Aviation Person-Hours: Commercial aviation person-hours were estimated by dividing the number of person-miles by 500 mph—the estimated average speed of a commercial aircraft. A sensitivity analysis of this estimate, using values ranging from 100 mph to 600 mph, did not affect commercial aviation's relative risk ranking in terms of fatalities per person-hour.

Exposure data that was estimated by the author are italicized in blue in Table 5; exposure data that could not be estimated (or identified in the literature) are listed as N/A.

4.6.1 Fatal Accidents per Vehicle-Trips

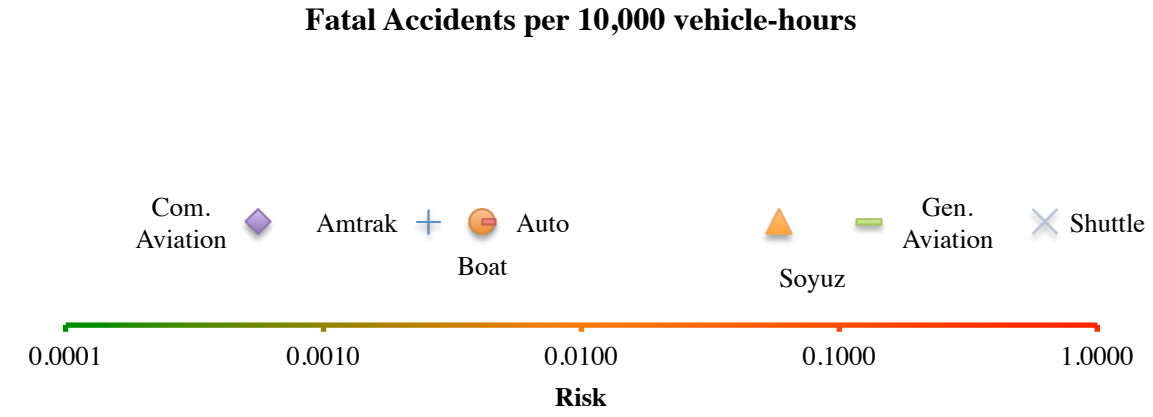


	Skydiving	SCUBA	Everest	Denali	Rainier	Auto	Amtrak	Boat	Gen. Aviation	Com. Aviation	Shuttle	Soyuz
Rank	N/A	N/A	N/A	N/A	N/A	6	5	4	3	7	2	1
Rate	N/A	N/A	N/A	N/A	N/A	$1x10^{-4}$	$2x10^{-3}$	$2x10^{-3}$	$8x10^{-3}$	$1x10^{-4}$	$1x10^{1}$	$2x10^{1}$

Figure 10: Number of fatal accidents per 1,000 vehicle-trips. Low risk activities appear on the left of the graph, while high risk activities appear on the right. Rankings ("Rank" in the table) are described in order of decreasing risk (1 = highest risk, 7 = lowest risk).

Space flight was measured to be riskier than all other terrestrial activities by several orders of magnitude on a fatal accidents per vehicle-trip basis (see Figure 10). Space Shuttle and Soyuz data, when extrapolated to 1,000 vehicle-trips, suffered 14.8 and 16.0 fatal accidents per vehicle-trip, respectively. In comparison, general aviation, the riskiest non-space activity on a fatal accident per vehicle-trip basis, experienced only 0.008 fatal accidents per 1,000 vehicle trips between 2003-2007—a rate that is statistically less than that experienced by the Space Shuttle and Soyuz (Chi-Square test, $p<0.01$).

4.6.2 Fatal Accidents per Vehicle-Hours

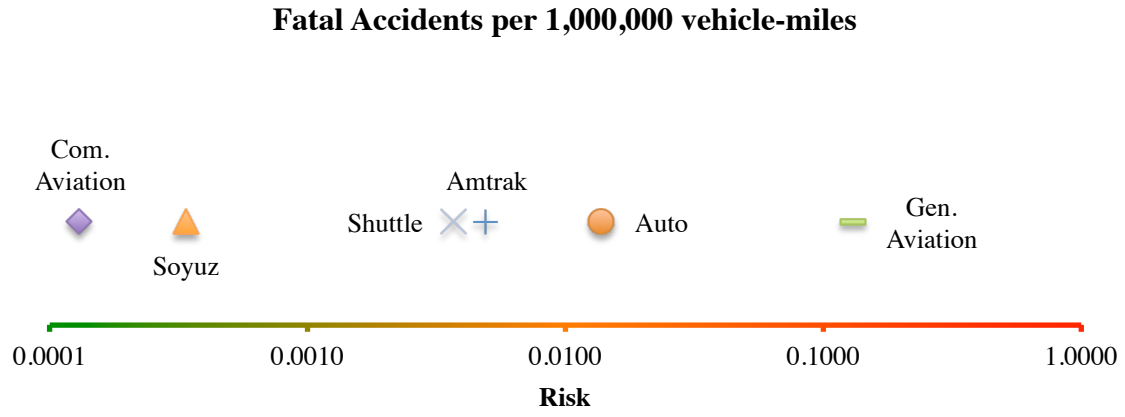


	Skydiving	SCUBA	Everest	Denali	Rainier	Auto	Amtrak	Boat	Gen. Aviation	Com. Aviation	Shuttle	Soyuz
Rank	N/A	N/A	N/A	N/A	N/A	5	6	4	2	7	1	3
Rate	N/A	N/A	N/A	N/A	N/A	$4x10^{-3}$	$3x10^{-3}$	$4x10^{-3}$	$1x10^{-1}$	$6x10^{-4}$	$6x10^{-1}$	$6x10^{-2}$

Figure 11: Number of fatal accidents per 10,000 vehicle hours. Low risk activities appear on the left of the graph, while high risk activities appear on the right. Rankings ("Rank" in the table) are described in order of decreasing risk (1 = highest risk, 7 = lowest risk).

Human space flight (as characterized by the Space Shuttle and Soyuz) was also identified as one of the riskiest activities when measured on a fatal accidents per vehicle-hour basis (see Figure 11). Of the 7 activities whose fatal accidents per vehicle-*hour* rates could be calculated, flights aboard the Space Shuttle were the riskiest (0.63 fatal accidents per 10,000 vehicle-hours). However, flights aboard general aviation aircraft (0.13 fatal accidents per 10,000 hours) were measured to be riskier than flights aboard the Soyuz spacecraft (0.06 fatal accidents per 10,000 vehicle-hours).

4.6.3 Fatal Accidents per Vehicle-Miles

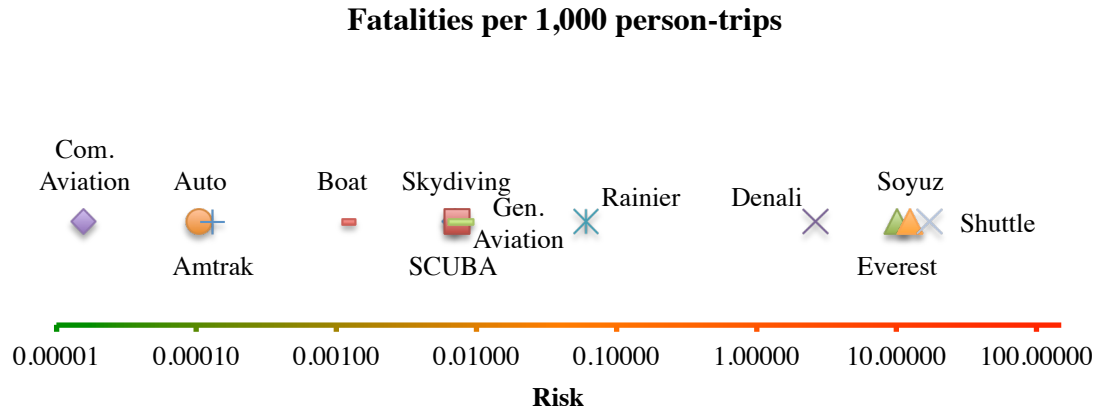


	Skydiving	SCUBA	Everest	Denali	Rainier	Auto	Amtrak	Boat	Gen. Aviation	Com. Aviation	Shuttle	Soyuz
Rank	N/A	N/A	N/A	N/A	N/A	2	3	N/A	1	6	4	5
Rate	N/A	N/A	N/A	N/A	N/A	$1x10^{-2}$	$5x10^{-3}$	N/A	$1x10^{-1}$	$1x10^{-4}$	$4x10^{-3}$	$3x10^{-4}$

Figure 12: Number of fatal accidents per 1,000,000 vehicle miles. Low risk activities appear on the left of the graph, while high risk activities appear on the right. Rankings ("Rank" in the table) are described in order of decreasing risk (1 = highest risk, 6 = lowest risk).

On a fatal accidents per vehicle-mile basis, human space flight was actually one of the *least* risky activities, experiencing only 0.0003 (Soyuz) and 0.0037 (Space Shuttle) fatal accidents per million vehicle-miles (see Figure 12). In contrast, automobiles (0.0137), Amtrak passenger rail (0.0049), and General Aviation (0.1299) experienced significantly more fatal accidents per vehicle-mile than both the Space Shuttle and Soyuz ($p<0.01$).

4.6.4 Fatalities per Person-Trips

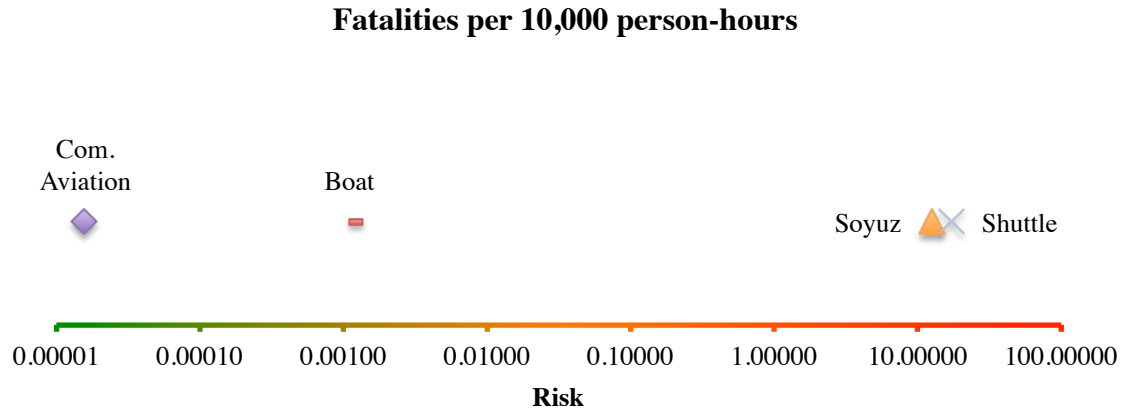


	Skydiving	SCUBA	Everest	Denali	Rainier	Auto	Amtrak	Boat	Gen. Aviation	Com. Aviation	Shuttle	Soyuz
Rank	7	8	3	4	5	11	10	9	6	12	1	2
Rate	$7x10^{-3}$	$7x10^{-3}$	$1x10^{1}$	$3x10^{0}$	$6x10^{-2}$	$1x10^{-4}$	$1x10^{-4}$	$1x10^{-3}$	$8x10^{-3}$	$2x10^{-5}$	$2x10^{1}$	$1x10^{1}$

Figure 13: Number of fatalities per 1,000 person-trips. Low risk activities appear on the left of the graph, while high risk activities appear on the right. Rankings ("Rank" in the table) are described in order of decreasing risk (1 = highest risk, 12 = lowest risk).

Both the Space Shuttle and Soyuz exhibited a higher number of fatalities per person-trip (17.1 and 12.6 per 1,000 person-trips, respectively) than any of the other reviewed activities (see Figure 13). However, the number of fatalities per person-trip for Mt. Everest (10.1 fatalities per 1,000 person-trips) was roughly comparable on an order of magnitude basis. In fact, statistical analysis indicates the fatalities per person-trip rates for Mt. Everest and Soyuz/Shuttle were not significantly different (Chi-square, $p<0.05$).

4.6.5 Fatalities per Person-Hours

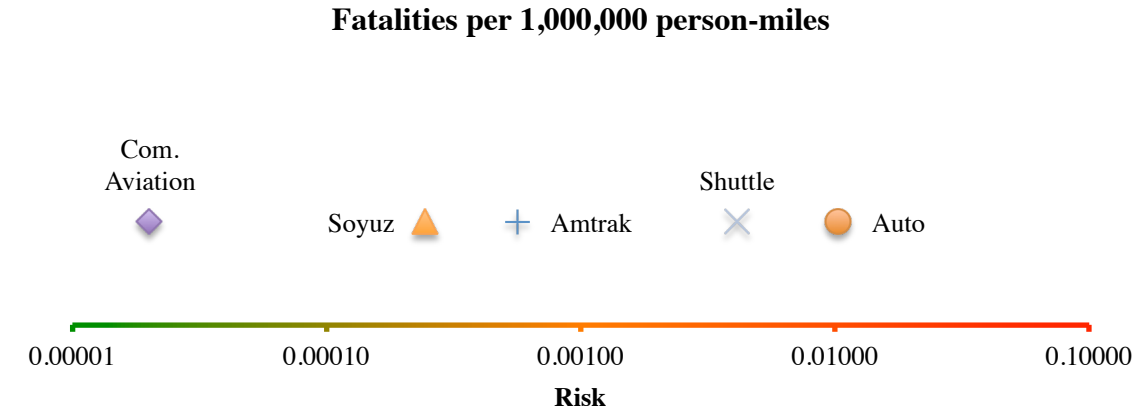


	Skydiving	SCUBA	Everest	Denali	Rainier	Auto	Amtrak	Boat	Gen. Aviation	Com. Aviation	Shuttle	Soyuz
Rank	N/A	N/A	N/A	N/A	N/A	N/A	N/A	3	N/A	4	1	2
Rate	N/A	N/A	N/A	N/A	N/A	N/A	N/A	$2x10^{-3}$	N/A	$1x10^{-4}$	$7x10^{-1}$	$4x10^{-2}$

Figure 14: Number of fatalities per 10,000 person-hours. Low risk activities appear on the left of the graph, while high risk activities appear on the right. Rankings ("Rank" in the table) are described in order of decreasing risk (1 = highest risk, 4 = lowest risk).

When measured on a *fatalities per person-hour* basis, human space flight (0.7012 fatalities per 10,000 person-hours for Space Shuttle, 0.0421 fatalities per 10,000 person-hours for Soyuz) was significantly riskier than either boating (0.0018 fatalities per 10,000 person-hours, Chi-square, $p<0.01$) or commercial aviation (0.0001 fatalities per 10,000 person-hours, Chi-square, $p<0.01$)—the only two other activities with comparable data reviewed in this thesis (see Figure 14).

4.6.6 Fatalities per Person-Miles



	Skydiving	SCUBA	Everest	Denali	Rainier	Auto	Amtrak	Boat	Gen. Aviation	Com. Aviation	Shuttle	Soyuz
Rank	N/A	N/A	N/A	N/A	N/A	1	3	N/A	N/A	5	2	4
Rate	N/A	N/A	N/A	N/A	N/A	$1x10^{-2}$	$6x10^{-4}$	N/A	N/A	$2x10^{-5}$	$4x10^{-3}$	$2x10^{-4}$

Figure 15: Number of fatalities per 1,000,000 person-miles. Low risk activities appear on the left of the graph, while high risk activities appear on the right. Rankings ("Rank" in the table) are described in order of decreasing risk (1 = highest risk, 5 = lowest risk).

Human space flight was calculated to be one of the riskiest activities when measured on a fatalities per person-mile basis (see Figure 15). However, it is worth noting that space flight was not the riskiest of all activities reviewed here, as the rate of automobile fatalities per person-mile (0.0103 fatalities per 1,000,000 person-miles) exceeded that of either the Space Shuttle (0.0041 fatalities per 1,000,000 person-miles) or Soyuz (0.0002 fatalities per 1,000,000 person-miles).

4.7 Discussion

By most reference units, human space flight was riskier than all other activities reviewed here. Flights on the Space Shuttle and Soyuz were significantly riskier than all other activities on a fatal accidents per vehicle-trip and a fatalities per person-hour basis, and one of the riskiest activities on a fatal accidents per vehicle-hour basis. Additionally, Space Shuttle and Soyuz flights were riskier than all other activities on a fatalities per person-trip basis (though not to a statistically significant degree).

However, when assessed on a per-mile basis (both fatal accidents per vehicle-mile and fatalities per person-mile), space flight was actually less risky than a number of other activities. For example, automobiles were identified to be significantly riskier than the Space Shuttle on both a fatal accidents per vehicle-mile and on a fatalities per person-mile basis.

These seemingly contradictory findings serve to demonstrate the limitations inherent to characterizing risk with a single reference unit. Yet the converse approach—characterizing risk with *a multitude* of reference units—may prove confusing, particularly if several reference units appear to present contradictory information.

One potential workaround is to present risk using only those reference units that directly relate to the activity's primary motive. Given that space flight is generally promoted as an experience, rather than a means of transportation, per-trip reference units (e.g., fatal accidents per vehicle-trip, fatalities per person-trip) may be the most appropriate means of characterizing risk for future spacecraft, at least at this stage of operations. NASA currently measures crewed spacecraft risk as a per-trip rate (and *only* as a per-trip rate), so there is also precedence to reporting risk in this manner. Consequently, risk will be reported using per-trip reference units (and only per-trip reference units) for the remainder of this thesis.

4.8 Chapter Summary

In this chapter, actuarial analysis was selected as the metric for measuring risk, as it avoids many of the complications associated with Probabilistic Risk Assessment. By selecting this metric, sub-question 2b, *how should risk be measured?* was answered.

Spacecraft risk was then measured using actuarial analysis, and compared to the risk of several different modes of transportation and adventure sport activities. This process characterized space flight risk in both an absolute and relative manner, thereby answering sub-question 2b, *how risky is space flight?*

4.9 Immediate Forward Work

The risk values identified in this chapter provide insight as to what level of risk is realistically achievable for future spacecraft. It also serves to identify how frequently each activity is pursued. Together these two concepts can be used to create a heuristic method for predicting spacecraft risk, as described in Chapter 5. From these predictions, "Safe Enough" can in part be determined.

4.10 Related Publications, Presentations, and Posters

Ocampo R. P. (2011). *Human Rating Space Systems: How Safe is Safe Enough?* Presented at ASGSB/AIAA Student Panel, Washington D.C.

Ocampo R. P. (2012, December). *History of Spacecraft Safety: U.S. Manned Space Program*. Presented at Bioastronautics Student Seminar, Boulder CO.

Klaus, D., Fanchiang, C., & Ocampo, R. (2012). Perspectives on Spacecraft Human-Rating. In *42nd International Conference on Environmental Systems* (p. 3419).

Ocampo, R. P., & Klaus, D. M. (2015, October). *Human Space Flight Safety*. Poster session presented at the FAA COE CST 5th Annual Technical Meeting, Washington, D.C.

Ocampo, R. P., & Klaus, D. M. (2016). *Comparing the Relative Risk of Space Flight to Terrestrial Modes of Transportation and Adventure Sport Activities*. Manuscript in preparation.

CHAPTER 5

DETERMINING SAFE ENOUGH

"It's a very sobering feeling to be up in space and realize that one's safety factor was determined by the lowest bidder on a government contract."

- Alan Shepard

5.1 Objective

Space flight will never be perfectly safe. Therefore, engineers must strive to design, develop, and operate spacecraft that are *safe enough*. As such, the two major objectives of this chapter are to:

1. Establish a means of distinguishing "Safe Enough" from "Not Safe Enough" that is consistent with the terms and framework defined in Chapter 3 and compliant with the risk metric selected in Chapter 4.
2. Develop a heuristic method, derived from risk and usage data collected in Chapter 4, which can predict spacecraft risk even in the absence of detailed spacecraft data. From these risk predictions, spacecraft can either be accepted as "Safe Enough" or rejected as "Not Safe Enough."

5.2 Background

5.2.1 Determining Safe Enough: Hazard Based Methods

Throughout its history, NASA has established several different methods for distinguishing "Safe Enough" from "Not Safe Enough." During the early days of human space flight, NASA relied on hazard analyses to delineate these two system states. Under this rubric, a spacecraft was considered "Safe Enough" if all hazards had a corresponding hazard control in place to eliminate or mitigate the hazard's likelihood or severity (NASA, 2011).

5.2.2 Determining Safe Enough: Requirements-Based Methods

Although hazard analyses are capable of identifying hazards associated with *individual* components (or subsystems), it is not well suited for evaluating hazards that arise from *systemic* failure (Dezfuli et al., 2011; ASAP, 2007). For this reason, NASA has shifted to a more holistic, requirements-based methodology for evaluating "Safe Enough" in recent years. Crewed vehicles owned or operated by NASA must now meet the requirements described in NPR 8705.2B, "Human-Rating Requirements for Space Systems" (or its derivative documents) to be considered programmatically acceptable for human spaceflight (NASA, 2008a; Klaus et al., 2014).

5.2.3 Rationale for new technique

A proscriptive "safe enough" methodology such as this can readily distinguish "Safe Enough" from "Not Safe Enough" based on the requirement set's verification state. However, this methodology also tends to bind spacecraft to a particular design "type" that may not always be optimal in terms of safety. Consider the case of Failure Tolerance (FT) requirements, which

are mainstays of most safety requirement sets (NASA, 2008a; NASA, 2015a; NASA, 2015b). FT requirements are ostensibly written to reduce risk, as Failure Tolerant spacecraft can—in theory—continue to function properly in the presence of one or more failures. However, Failure Tolerance can also add complex and unpredictable failure modes to the spacecraft—which in turn can lead to an overall *increase* in risk in certain cases.

For example, the addition of a redundant depressurization valve on Soyuz 11 was intended to protect the crew against pressure equalization failures during reentry and landing, but ultimately contributed to the vehicle's catastrophic depressurization and Loss of Crew (LOC) when it inadvertently opened in flight, allowing the cabin atmosphere to vent overboard (Ocampo, 2014).

In instances like this, a spacecraft identified by a requirements-based methodology as "Safe Enough" (e.g., all requirements verified, including all FT requirements) would actually be *riskier* than a spacecraft identified as "Not Safe Enough" (e.g., not all requirements verified, not fully Failure Tolerant)—an outcome that demonstrates the potential fallibility of requirements-based "Safe Enough" rubrics.

Given the limitations inherent to both hazard-based and requirements-based "Safe Enough" methodologies, this thesis proposes an alternative method for distinguishing "Safe Enough" spacecraft from "Not Safe Enough" spacecraft.

5.3 Redefining Safe Enough

The definitions and framework established in Chapter 3 (as well as the actuarial risk metric selected in Chapter 4) logically suggest that "safe enough" spacecraft can be delineated as follows: spacecraft that exhibit risk values that are *less than or equal to* an established "risk

threshold" can be considered "Safe Enough"; conversely, spacecraft that exhibit risk values that are *greater* than an established risk threshold can be rejected as "Not Safe Enough" (see Figure 16).

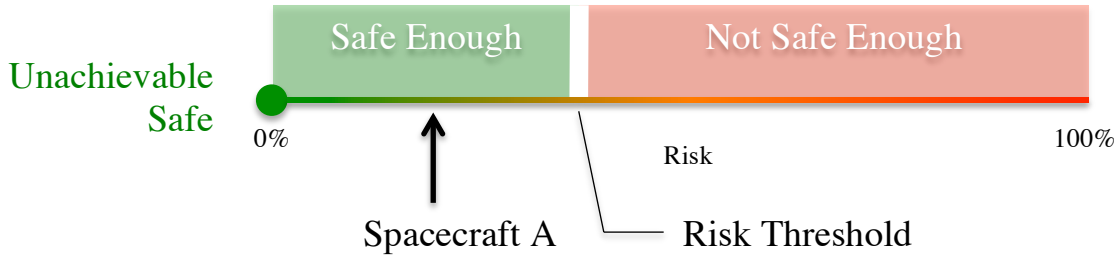


Figure 16: Spacecraft with risk values that are less than or equal to an established risk threshold can be considered "Safe Enough." Spacecraft with risk values that are not less than the risk threshold can be rejected as "Not Safe Enough."

5.4 Challenges of Determining Risk Thresholds

Given this rubric, determining whether a spacecraft is "Safe Enough" is mathematically trivial. *Establishing an appropriate risk threshold value*, however, constitutes a far greater challenge. Risk thresholds values must balance what is maximally *acceptable* from a policy standpoint with what is minimally *achievable* given technical, budget, and schedule constraints (Ocampo & Klaus, 2016a; ASAP, 2014) (see Figure 17).

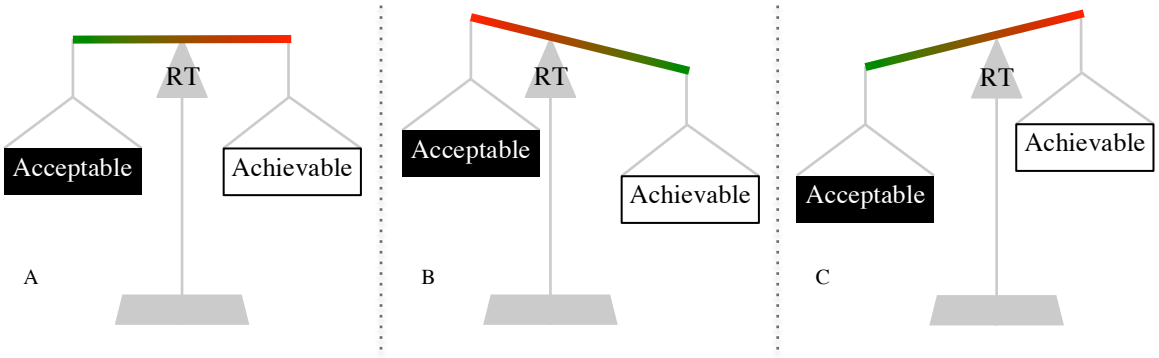


Figure 17: A) Risk thresholds (RT) must balance what is acceptable from a programmatic standpoint with what is achievable given technical, budget, and schedule constraints. B) Risk thresholds that are set too low (left side of the scale) will be highly acceptable (*high acceptability*) but difficult to achieve (*low achievability*); C) risk thresholds that are set too high (right side of the scale) will be relatively easy to achieve (*high achievability*) but difficult to accept (*low acceptability*).

This "balanced" value must be established during the initial stages of a program's lifecycle, when it can most effectively guide the development of operational and design requirements. Yet determining achievable risk[8] requires a set of mature risk analysis products—*which cannot be developed without firm operational design requirements in place* (Turner et al., 2005; NASA, 2011). This web of cyclic relationships helps to explain why risk thresholds are generally developed in an iterative fashion, as depicted in Figure 18.

[8] As mentioned above and depicted in Figure 18, *achievable* risk is only half of the risk threshold equation; in order for a balanced risk threshold to be determined, a level of *acceptable* risk must also be established. Calculating acceptable risk is "far from [a] straightforward" process, as anticipated mission benefits must be shown to demonstrably outweigh the potential for mishap (ASAP, 2014; ASAP, 2015). The process and implications of determining acceptable risk, however, are not described further in this chapter, as they are discussed in detail in the Conclusions chapter.

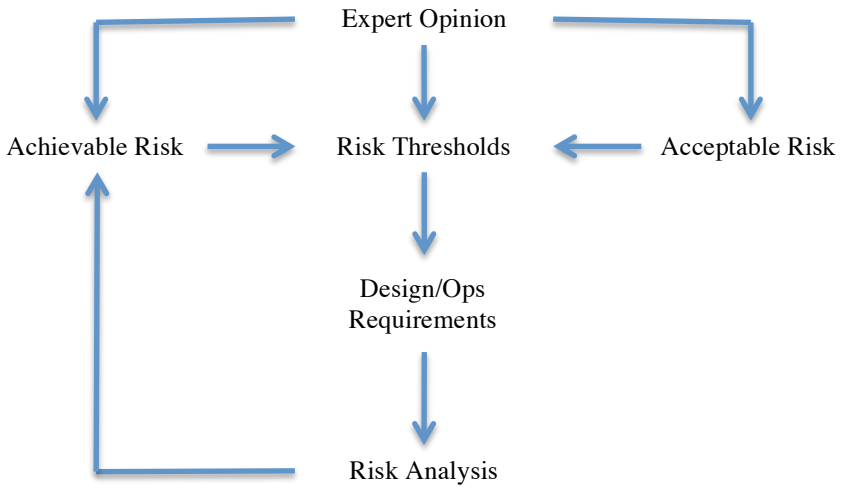


Figure 18: Determining a risk threshold requires, in part, an understanding of the system's level of achievable risk. Calculating achievable risk requires mature risks analysis products and design/operational requirements, which are difficult (if not impossible) to effectively develop without a risk threshold in place. This "chicken or egg" process tends to lead to iteratively developed risk thresholds, which can be difficult to initiate as there is no obvious "start" point.

Given the cyclic, iterative nature of this process, crewed space programs tend to rely on expert opinion when setting risk thresholds, at least initially (ASAP, 2014) (Figure 18). As knowledge and experience with the system are acquired, however, quantitative data can be blended with expert opinion to form increasingly accurate predictions of achievable risk.

These improving predictions can then be used as rationale for resetting the risk threshold, as necessary. The Constellation Program, for example, was required to meet a risk threshold of 1/1000 for its mission to the International Space Station (ISS), but this value was increased to 1/270 when NASA realized the Orion and Ares spacecraft could not meet the more demanding requirement (ASAP, 2012). In a similar fashion, NASA's Commercial Crew Program (CCP) increased its risk threshold from 1/270 to 1/200 when analysis indicated the risk associated with Micrometeoroid and Orbital Debris (MMOD) would preclude CCP spacecraft from meeting the original requirement (ASAP, 2016).

This tendency to incrementally "relax" risk thresholds over the course of program development can lead to spacecraft with risk levels that are technically achievable. However, this approach can also inadvertently lead to spacecraft that are *programmatically unacceptable* from a risk perspective (ASAP, 2015) (Figure 17**c**). Such designs rarely fly (or rarely fly for long); the Space Shuttle, for example, was retired in 2003 because the vehicle's increasing risk was no longer considered tenable (CAIB, 2003).

5.5 Goals

Early *and* accurate predictions of achievable risk are therefore critical to a program's viability, as they serve to influence (and in turn, be influenced by) the risk threshold (see Figure 18). If risk is predicted to be acceptably low, the program can continue forward; if risk is predicted to be unacceptably high, the program can be restructured (or cancelled) before significant resources are committed to a specific design.

This chapter presents a novel method for predicting achievable risk that is modeled upon the risk progression rates (seemingly) inherent to complex systems. This risk heuristic is independent of any specific design or program, and as such, does not share many of the limitations inherent to most risk prediction methods described above.

5.6 Rationale

While space flight risk has a tendency to *increase* during program development (ASAP, 2012; ASAP, 2016), it has a tendency to *decrease,* in a quasi-predictable fashion, over the course of a program's operational lifetime. This reduction in risk can be seen in both probabilistic and actuarial measures of spacecraft risk. Space Shuttle mean probability of Loss of Crew (LOC)

values declined logarithmically over the course of the program, from 1 in 12 at STS-1 to 1 in 90 at STS-133 (Hamlin et al., 2011) (see Figure 19). Launch vehicle cumulative failure rates also show a similar decrease with usage, typically in a manner that can be fitted to a logarithmic function (see **Figure 20**).

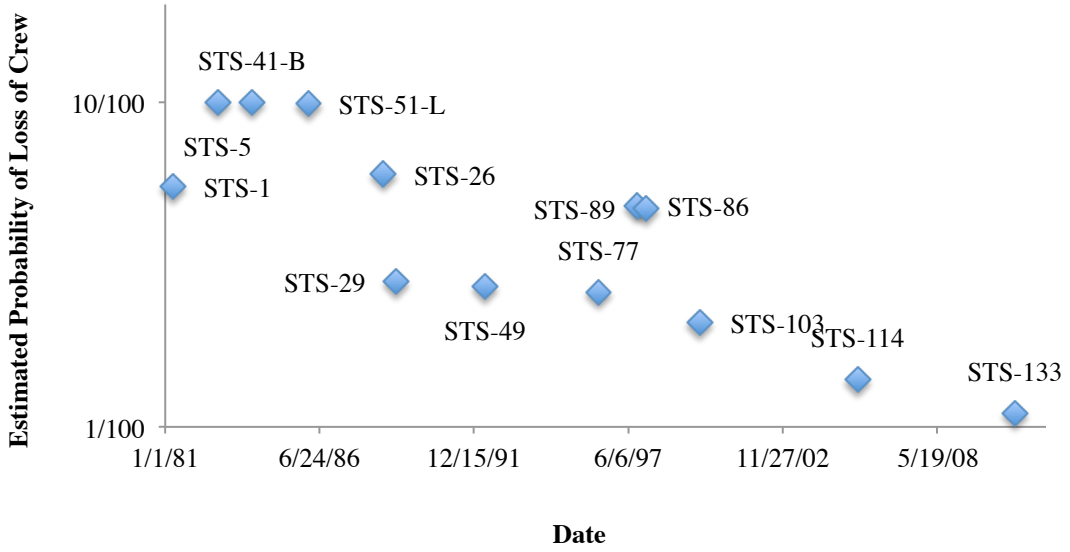


Figure 19: Space Shuttle estimated Probability of Loss of Crew over time (PRA data). Data derived from Hamlin et al., 2011.

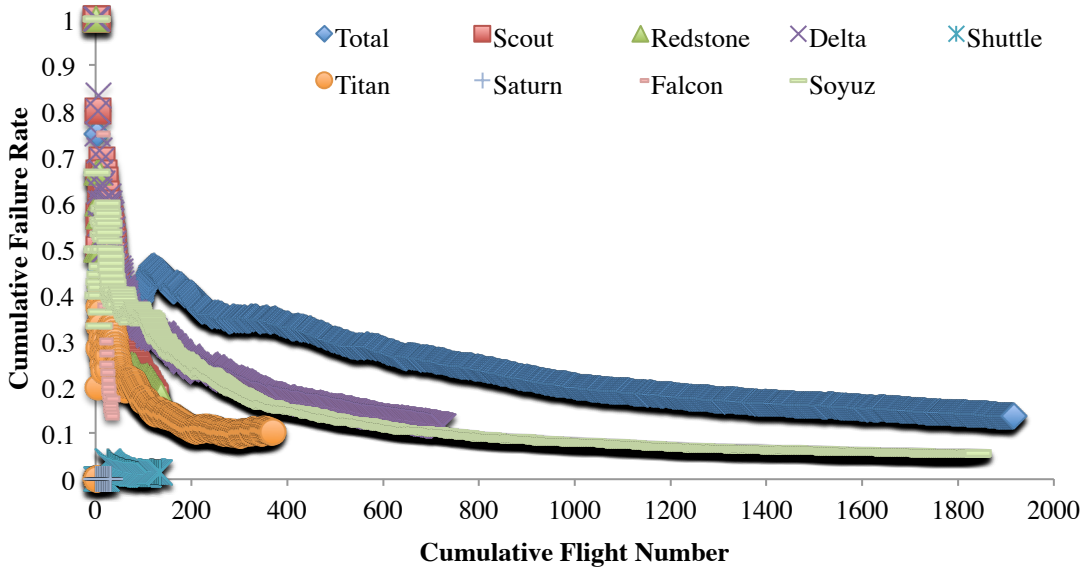


Figure 20: Cumulative Launch Vehicle Failure Rate vs. Flight Number. As the number of flights increases, the cumulative failure rate decreases in a logarithmic fashion.

These risk progression trends are not limited to space systems. The total number of fatal automobile accidents (per billion miles of vehicle travel) also declined logarithmically from 1994 to 2013, decreasing 34% (Figure 21) (NHTSA, 2016). The total number of fatal general and commercial aviation accidents (per thousand hours) also exhibited a similar logarithmic decline (from 1970 to 2010), decreasing 57% and 95%, respectively (BTS 2016b, BTS 2016c) (Figure 22).

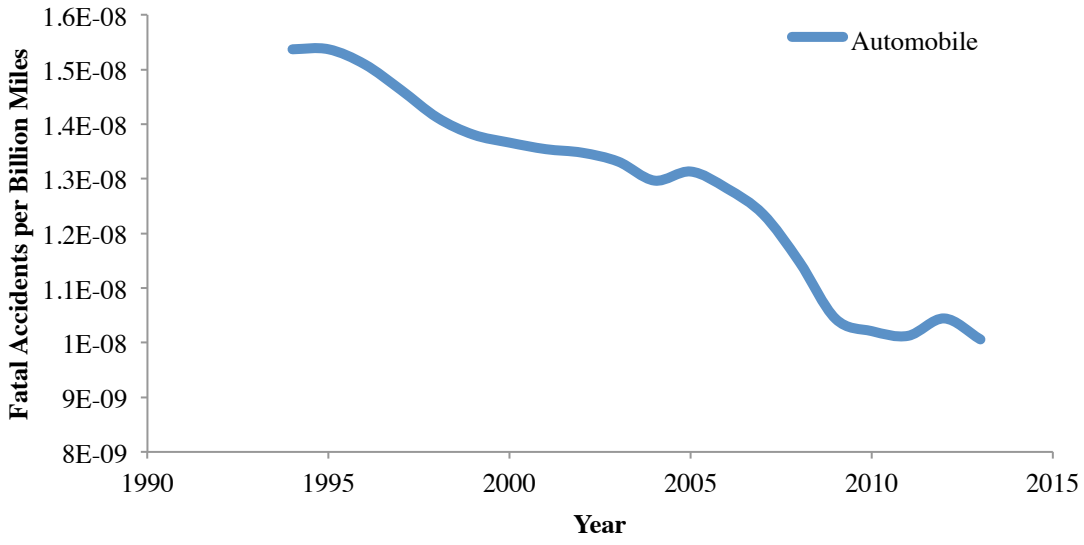


Figure 21: Automobile risk reduction rates for a 20 year time period ranging from 1994-2013 (NHTSA, 2016).

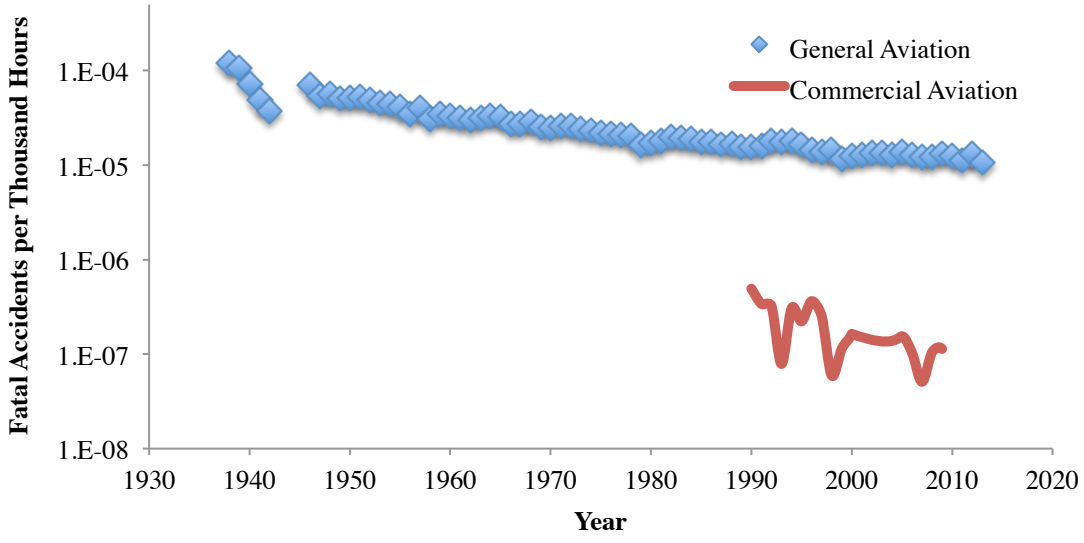


Figure 22: Risk reduction rates for a) automobiles and b) general and commercial aviation (BTS 2016b, BTS 2016c).

This relationship between risk and usage is even present in non-transportation activities. Mountaineering fatality rates (e.g., fatalities per total number of participants) on Mt. Everest, Denali, and Mt. Rainier, for example, each exhibited a logarithmic decline over the last 50 years (Salisbury & Hawley, 2011; Waterman, 1991; McIntosh et al., 2008; NPS, 2008-2015; NPS, 2016; NPS, 2010) (see **Figure 23**).

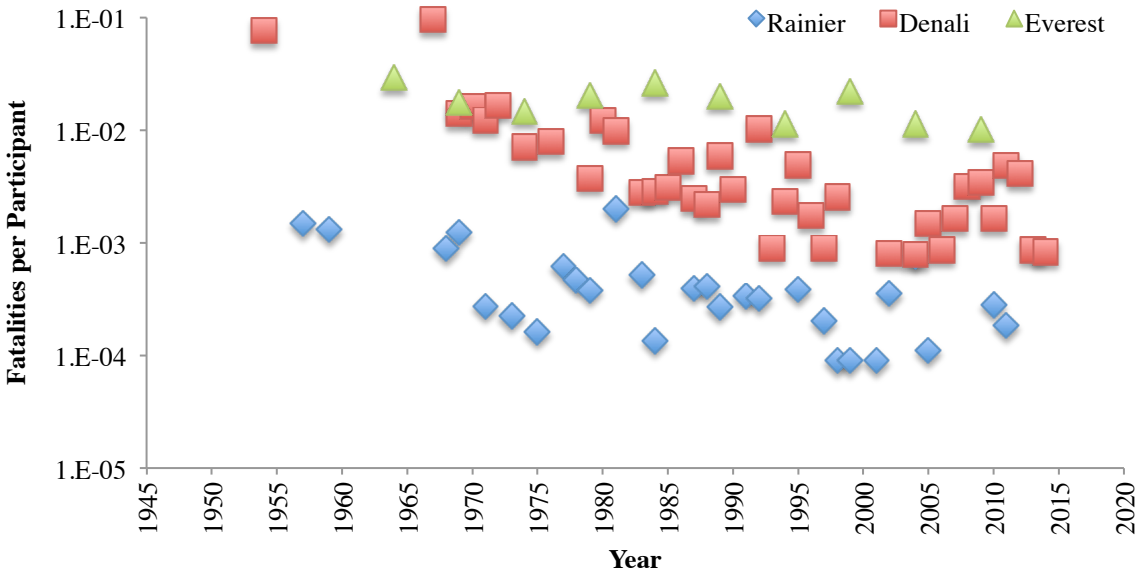


Figure 23: Mountaineering fatality rates vs. year for Mt. Everest, Denali, and Mt. Rainier.

5.7 Methodology

This inverse relationship between risk and usage, which appears across multiple unrelated activities, suggests the two variables may be highly correlative. If this relationship is true, it may be possible to coarsely predict achievable space flight risk knowing only the anticipated flight rate of the spacecraft (and vice versa).

Given the potential benefits of such a predictive technique, risk and usage data for each of the different activities reviewed in Chapter 4 were correlated and fitted to an equation using logarithmic regression analysis. (For a detailed description of how these metrics were measured or estimated, see Chapter 4.)

5.8 Results

Log-log plots of risk and usage from both a person-centric (Figure 24) and vehicle-centric (Figure 25) perspective are presented below.

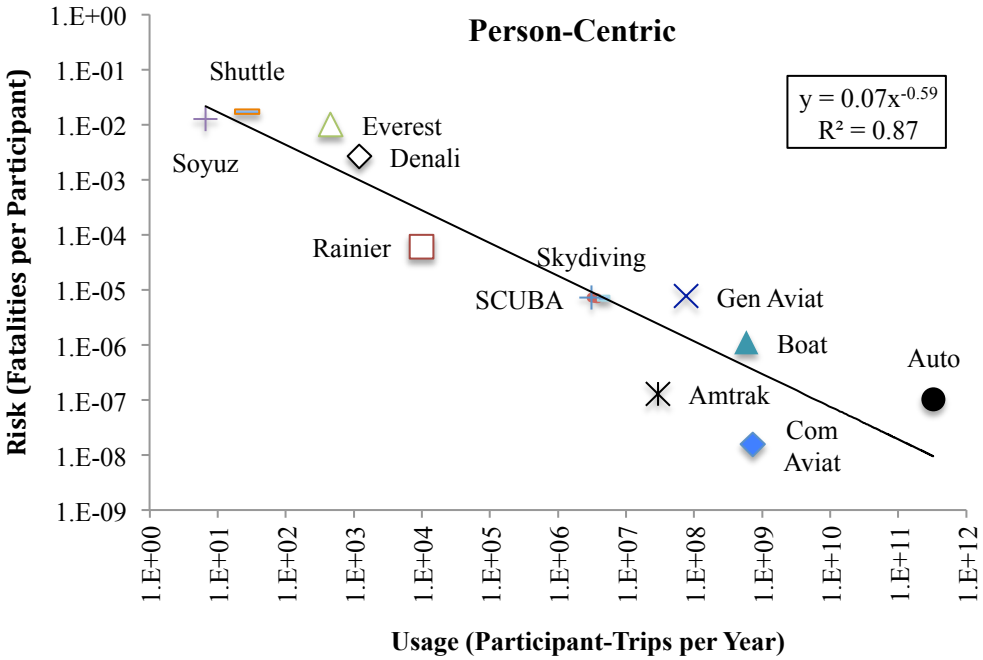


Figure 24: Log-log plots of risk (fatalities per person-trip) and usage (participants per year) from a person-centric perspective.

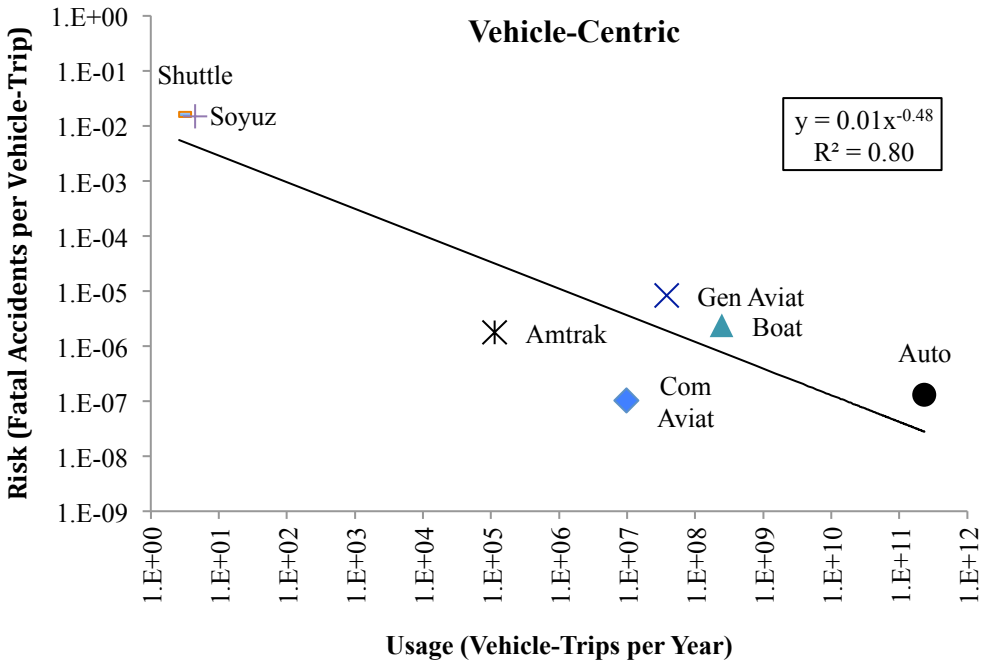


Figure 25: Log-log plots of risk (fatal accidents per vehicle-trip) and usage (vehicle-trips per year) from a vehicle-centric perspective.

From a *person-centric* perspective, risk and usage were highly correlated (r = -0.93) (Figure 24). Despite fewer data points, risk and usage also exhibited a strong correlation when assessed on a *vehicle-centric* basis (r = -0.90)(Figure 25)[9]. Both correlations were statistically significant, with p-values less than 0.01 (1.10×10^{-5} and 6.33×10^{-3} for *person-centric* and *vehicle-centric* data, respectively).

A regression analysis was performed on both person-centric and vehicle-centric data, and the resulting logarithmic functions were plotted in Figure 24 and Figure 25. These functions

[9] As described in Chapter 4, vehicle-centric risk metrics cannot be used to measure adventure sport activities, as vehicles are generally not involved in these activities. This explains why there are fewer data points.

exhibited high goodness of fit, with R^2 values equal to 0.87 and 0.80 for person-centric and vehicle-centric data, respectively.

5.9 Discussion

Based on this analysis (and this analysis alone), the relationship between risk and usage can be labeled highly correlative, but *not necessarily* causal. There is, however, strong *qualitative* evidence to suggest the two variables may indeed share a cause and effect relationship, as described below.

5.9.1 Increasing Usage, Decreasing Risk

An increase in usage provides an environment where lessons can be rapidly learned and assimilated, both at the hardware and operations level. This in turn can lead to an overall reduction in risk. Project Gemini's success, for example, can be partially attributed to the rapid buildup of space flights that occurred between 1965-1966 (Hacker & Grimwood, 1977).

An increase in usage can also serve as an impetus for infrastructure improvements, ranging from the addition of traffic lights at busy intersections, to the construction of air traffic control towers at congested airports, to the use of fixed lines on heavily climbed mountains, such as Mt. Everest. These improvements in turn can lead to commensurate decreases in risk.

However, this relationship between increased usage and decreased risk appears to hold true only when there are sufficient resources available to incorporate new data (e.g., "lessons learned"). Prior to the *Challenger* accident, the Space Shuttle Program was "approaching a state of saturation in which no more flights could be accommodated" (U.S. House of Representatives,

1986, p. 121); in this environment, lessons learned from one mission could not adequately be incorporated into the next (Rogers et al., 1986).

5.9.2 Decreasing Usage, Increasing Risk

A decrease in usage can lead to the atrophy of technical skills and expertise; this in turn can lead to an overall increase in risk. In their 2009 review of United States Human Space Flight Plans, the Augustine Committee stated that one of the benefits to flying the Space Shuttle beyond its scheduled retirement date, at a "minimum safe flight rate," was the preservation of workforce and skills that "enable the U.S. to enjoy a robust human spaceflight program" (Augustine, 2009, p. 51). In a similar vein, ASAP has expressed concerns regarding the infrequent flight rate of the Space Launch System (SLS), stating that the proposed flight manifest may lead to an overall increase in risk due to "personnel loss and fading memories" (ASAP, 2016, p. 7).

5.9.3 Increasing Risk, Decreasing Usage

For certain activities, the relationship between risk and usage, if causal, may also be bidirectional. In other words, usage may affect risk (as noted above), but *risk may also affect usage*. For instance, participants may be less inclined to perform an activity or use a mode of transportation if there is an increase in the perceived risk of the system. After the 9/11 terrorist attacks, for example, commercial airlines experienced a 30% reduction in passengers in the months which followed (BTS, 2005).

5.9.4 Decreasing Risk, Increasing Usage

A decrease in risk can also readily lead to an increase in usage. The use of improved weather forecasts, better technical equipment, and superior navigational tools has contributed to an increase in usage in activities ranging from commercial aviation to mountaineering.

5.10 Predictive Capabilities

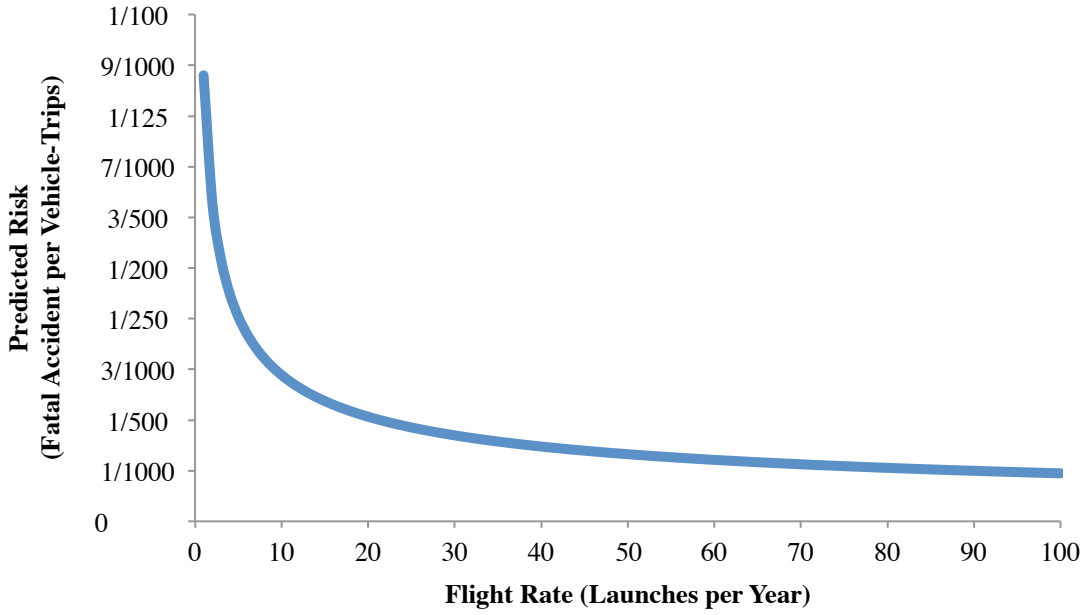
This relationship between risk and usage, though *not definitively* causal, is highly correlative; as such, it may prove useful as a coarse predictor of achievable risk, particularly during the early stages of program development when quantitative data is not readily available to support standard risk prediction techniques, such as PRA.

Using the vehicle-centric regression equation, achievable risk was predicted for a range of space flight rates, from 1 to 100 launches per year (see Figure 26). This risk heuristic indicates that a flight rate of 3-4 flights per year roughly corresponds to a 1/200 risk value—which is the *probabilistic*[10] value currently required for NASA's commercial crew program (NASA, 2015a). Tripling the flight rate to 10 launches a year further reduces risk by roughly 50%, to 1/329 (see Table in Figure 26).

However, the data seems to indicate that further reductions in achievable risk may require a dramatic increase in the flight rate, given the logarithmic relationship between risk and usage. A risk value of 1/1000 (as requested by the astronaut office for the ascent portion of flight in 2004 [Rominger, 2004]), equates to a flight rate of roughly 100 flights per year (see Table in Figure 26). This flight rate exceeds the maximum flight rate for the Space Shuttle (9 launches in

[10] Probabilistic and actuarial values (which are the types of values predicted in this analysis) can differ, sometimes drastically as described in Chapter 4; however, high-fidelity probabilistic estimates should ultimately converge on actuarial values over time.

1985) by a factor of ten (Ocampo, 2015) and even eclipses the optimistic prediction of a weekly flight rate predicted for the Space Shuttle in the 1970s (Theurer, 1976).



Flight Rate	1	2	3	4	5	6	7	8	9	10	20	40	60	80	100
Risk (1 in x)	114	159	193	222	248	270	291	311	329	346	484	678	824	948	1056

Figure 26: Predicted Risk vs. Flight Rate.

5.11 Discussion

This method of predicting achievable risk identified here is based solely on correlative analysis, and should not be used to substantiate an increase (or decrease) in flight rates in the absence of additional qualitative and quantitative evidence. Nevertheless, given the strong relationship between risk and usage, this risk heuristic may prove useful as a tool for coarsely

measuring risk, particularly during the early stages of vehicle development, when quantitative data may be limited or immature.

This method is not intended to replace vehicle-specific risk analysis techniques, such as PRA, but to work in tandem with them across varying stages of design and operations. During the initial phases of a program, this risk heuristic can provide valuable first-order approximations of risk; as data matures and becomes more readily available, PRA and other quantitative technique can be used to fine-tune the initial risk estimate (see Figure 27).

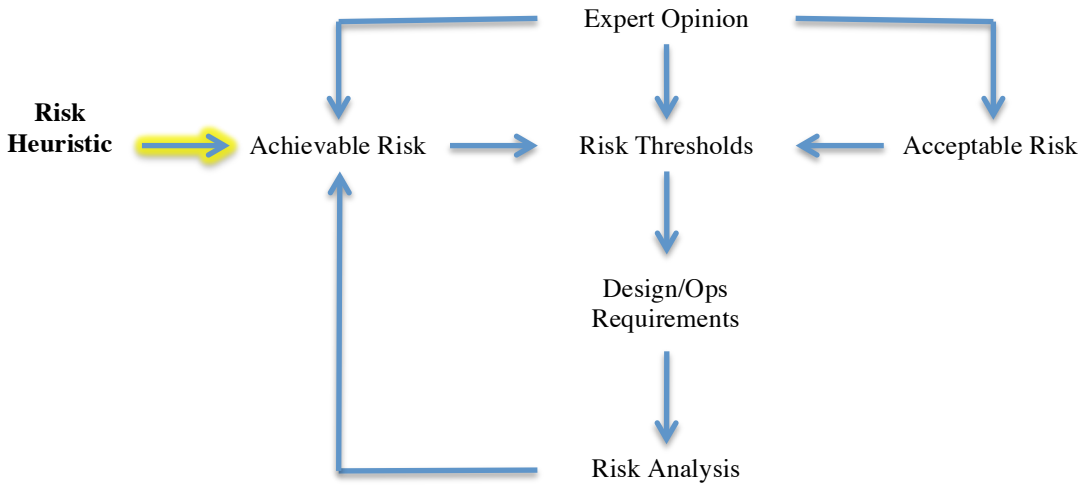


Figure 27: The risk prediction method presented here is meant to work in tandem with PRA and other risk analysis techniques to identify a risk threshold that is achievable.

Together these two methods can be used to predict achievable risk *throughout* a program's development and operations. If achievable risk is calculated to be acceptably low, the system can be considered "Safe Enough"—until proven otherwise. If achievable risk is calculated to be unacceptably high (e.g. "Not Safe Enough"), changes to the system's flight rate, design, and/or operations can be used to potentially achieve "Safe Enough."

5.12 Chapter Summary

This chapter established a means of distinguishing "Safe Enough" from "Not Safe Enough" that is consistent with the terms and framework defined in Chapter 3 and compliant with the risk metric selected in Chapter 4. This method of distinguishing risk serves to answer sub-questions 3a, *what does it meant to be "safe enough?"*

A method for predicting achievable risk is also established. This method is independent of any specific design or program, and as such, can be used to predict achievable levels of risk in the absence of detailed system data. This method serves to answer sub-question 3b, *what is the minimum level of risk that can be achieved?*

5.13 Related Publications, Presentations, and Posters

Ocampo, R. P. (2015). The Space Shuttle's Commercial Potential: A Retrospective Analysis. In *AIAA SPACE 2015 Conference and Exposition* (p. 4614).

Ocampo, R. P., & Klaus, D. M. (2015, October). *Human Space Flight Safety*. Poster session presented at the FAA COE CST 5th Annual Technical Meeting, Washington, D.C.

Ocampo, R. P., & Klaus, D. M. (2016). *A Novel Method for Predicting Achievable Risk in Human Space Flight*. Manuscript in preparation.

CHAPTER 6

CONCLUSIONS

"If you are looking for perfect safety, you will do well to sit on a fence and watch the birds; but if you really wish to learn, you must mount a machine and become acquainted with its tricks by actual trial."

- Wilbur Wright

6.1 Overview

The primary objective of this thesis was to resolve—or at least, *attempt to resolve*— the question "how safe is safe enough in human space flight?" This question was broken down into three steps, which served to answer seven sub-questions:

Step 1: Define Terminology

1a) What does it mean to be "safe?"

1b) What does it mean to be "unsafe?"

1c) What does "risk" mean?

Step 2: Characterize Risk

2a) How should risk be measured?

2b) How risky is space flight?

Step 3: Determine "Safe Enough"

3a) What does it mean to be "safe enough?"

3b) What minimum level of risk is achievable?

6.2 Summary

In the process of answering these seven sub-questions, 3 complications—*uncertainty in terminology*, *uncertainty in the choice of metrics*, and *uncertainty in the measurement itself*—were mitigated to the maximum extent practicable. In addition, a subjective decision relating to the *choice of metric* was made and logically supported. Complication 5, *subjectivity in the acceptance of the metric*, is discussed at the end of this chapter.

What does it mean to be safe? What does it mean to be unsafe? What does risk mean?

This thesis defined a "Safe" spacecraft as one that is free from all catastrophic hazards. Given that no real-world spacecraft can be free from *all* catastrophic hazards, this state was deemed unachievable in practice. Instead, all real-world spacecraft are by definition inherently "Unsafe." The *degree* to which they are "Unsafe" (e.g., their degree of risk) is directly related to the probability they will experience one or more catastrophic hazards.

How should risk be measured? How risky is space flight?

This probability can be quantified using either Probabilistic Risk Analysis (PRA) or actuarial analysis. Actuarial analysis was selected as the metric for measuring risk in this thesis because it produces measurements of risk with zero uncertainty. In addition, this metric has a large set of non-proprietary data readily available for analysis.

Actuarial analysis of the combined Space Shuttle and Soyuz safety records indicate that the historic risk of space flight is 1 fatal accident per 65 flights and 1 fatality per 63 participants. Given these safety records (and using these reference units), comparative analysis suggests that

space flight has historically been riskier than most transportation modes and adventure sport activities.

What does it mean to be safe enough?

Given the definitions of "safe," "unsafe," and "risk," a spacecraft can be considered "Safe Enough" if it has a risk value that is less than or equal to an established risk threshold. If the risk value is greater than this established risk threshold, it can be rejected as "Not Safe Enough."

What is the minimum level of risk that can be achieved?

The level of risk that can be achieved by a given spacecraft appears to be inversely related to its usage. A spacecraft that flies twice a year is predicted to have 1 fatal accident every 159 flights; a spacecraft that flies 100 times a year is predicted to have 1 fatal accident every 1056 flights.

This risk heuristic provides insight as to how risk can potentially be reduced. The strong correlation between risk and usage suggests that risk can be mitigated not by flying less often, *but by flying more often*[11].

6.3 Closing Thoughts

Determining whether the predicted risk values (or any risk value) should be accepted as "Safe Enough" is beyond the scope of this thesis. The risk framework and risk prediction

---

[11] That being said, arbitrarily increasing the launch rate without an attendant increase in resources and infrastructure can (and likely, will) lead to an *increase* in risk, as it did in the years preceding the Challenger accident (Rogers et al., 1986). However, all things being equal, the evidence presented here strongly suggests that increasing flight rates may be the key to reducing space flight risk in the long term, thereby achieving "Safe Enough."

heuristic presented here are tools that can be used to measure, compare, predict, and potentially mitigate risk; however, the actual decision to *accept* risk is one that must be made ***not*** by the author, but by society at large. To Gus Grissom and his fellow astronauts, the "conquest of space [was] worth the risk of life." Now we must ask: is it still?

CHAPTER 7

FORWARD WORK

""All this will not be finished in the first one hundred days. Nor will it be finished in the first one thousand days . . .nor even perhaps in our lifetime on this planet. But let us begin."

- President John F. Kennedy

The primary limitation of the risk prediction method presented here is that it relies on spacecraft data that in some cases is over 40 years old. This data may no longer accurately represent current space flight risk.

Soyuz suffered 2 fatal accidents and 4 fatalities over the course of its first 125 flights (dating to March 2015). However, both accidents (and all 4 fatalities) occurred early in the program's history (1967 and 1971). Given that the hazards that led to these fatal accidents have apparently been mitigated (as evidenced by the fact that no parachute or equalization valve has failed in the past 40 years), Soyuz safety records may not accurately reflect *current* Soyuz risk. If only the spacecraft safety records from the last 5 years of operations are included in the analysis (as was the protocol for all other activities reviewed here), then both Soyuz and Space Shuttle safety records would be perfect (0 fatal accidents and 0 fatalities over the course of 21 flights for both Soyuz and Shuttle).

To both validate and further refine the risk prediction method presented here, the correlative data upon which this method is based on must continuously be updated. This holds particularly true for nascent commercial spacecraft, as the safety records from their early

operations are unlikely to represent the mature risk of the spacecraft. To a less frequent extent, the safety records of terrestrial modes of transportation and adventure sport activities must also be updated to account for significant changes in risk and/or usage (e.g., "driverless" cars, high-speed passenger rail, next generation supersonic transport).

New data sources may also serve as a mechanism for validating and refining the risk prediction method. For example, submarine safety records, which were not available for this analysis, may prove useful in assessing how risk and usage relate when activity use is "medium" (i.e., the activity is performed more often than space flight, but less often than other, more common forms of terrestrial transportation).

In addition, uncrewed launch vehicle (LV) safety records (which were not included in the original analysis due to the fact that their use, by definition, involves no fatal risk) may provide relevant data that can improve the fidelity of the risk prediction method. In fact, when uncrewed LV safety records from 1953-2012 were included in the vehicle-centric correlation analysis, the R^2 regression analysis value actually improved from 0.80 to 0.85 (see Figure 28).

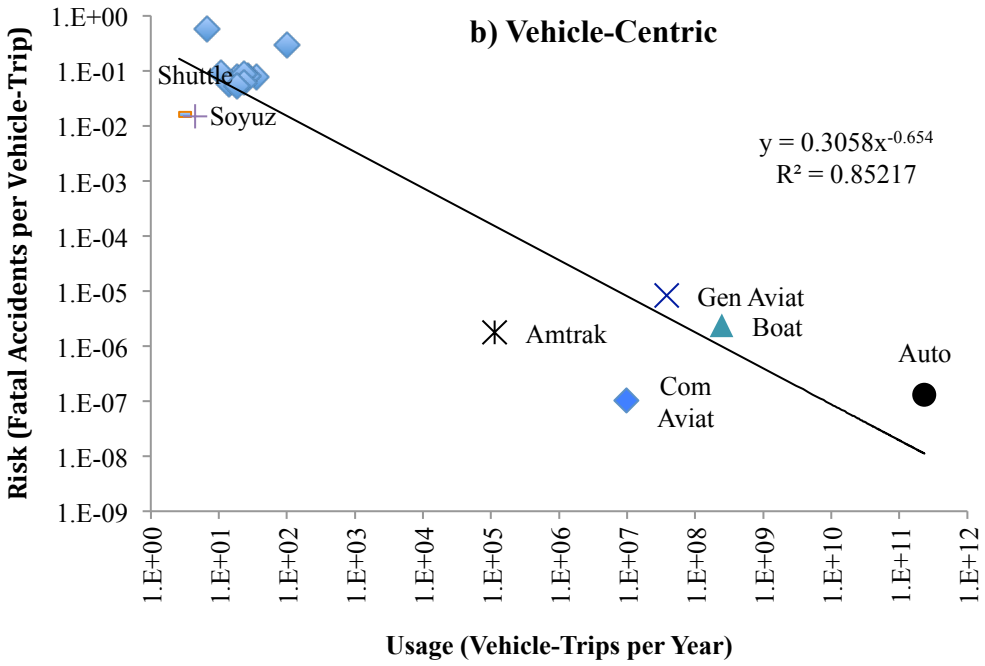


Figure 28: Log-log plots of risk (fatal accidents per vehicle-trip) and usage (vehicle-trips per year) from a vehicle-centric perspective. This plot includes uncrewed launch vehicle safety records (blue diamonds) from 1953-2012, binned into 5 year periods.

Lastly, vehicle-centric and person-centric safety records from previous decades may also prove valuable in assessing the correlation between risk and usage. General aviation safety records extend all the way to the 1940s, and may prove useful in further refining and validating the risk prediction method.

Perhaps one day, humans will be able to model risk with nearly perfect fidelity and reduce risk to nearly zero. Until then, however, the best method of predicting and reducing risk may be to fly in the face of it; indeed, as Wilbur Wright once said, "no bird soars in a calm."

SUMMARY OF PUBLICATIONS, PRESENTATIONS, AND POSTERS

Journal Articles

Ocampo, R. P., & Klaus, D. M. (2013). A Review of Spacecraft Safety: From Vostok to the International Space Station. *New Space*, *1*(2), 73-80.

Ocampo, R. P., & Klaus, D. M. (2016). A Quantitative Framework for Defining "How Safe is Safe Enough" in Crewed Spacecraft. *New Space*. Manuscript in press.

Ocampo, R. P., & Klaus, D. M. (2016). Comparing the Relative Risk of Space Flight to Terrestrial Modes of Transportation and Adventure Sport Activities. Manuscript in preparation.

Ocampo, R. P., & Klaus, D. M. (2016). *A Novel Method for Predicting Achievable Risk in Human Space Flight*. Manuscript in preparation.

Conference Papers

Klaus, D., Fanchiang, C., & Ocampo, R. (2012). Perspectives on Spacecraft Human-Rating. In *42nd International Conference on Environmental Systems* (p. 3419).

Klaus, D., Ocampo, R., & Fanchiang, C. (2014, March). Spacecraft human-rating: Historical overview and implementation considerations. In *Aerospace Conference, 2014 IEEE* (pp. 1-7). IEEE.

Ocampo, R. P. (2014). Limitations of Spacecraft Redundancy: A Case Study Analysis. 44th International Conference on Environmental Systems.

Ocampo, R. P., Herbert, B., & Turner, J. (2015). Linking Spacecraft Hazard Controls with System Design Requirements: General Considerations and Complications. In *AIAA SPACE 2015 Conference and Exposition* (p. 4637).

Ocampo, R. P. (2015). The Space Shuttle's Commercial Potential: A Retrospective Analysis. In *AIAA SPACE 2015 Conference and Exposition* (p. 4614).

Conference Presentations

Ocampo R. P. (2011, September). *Human Rating Space Systems: How Safe is Safe Enough?* Presented at ASGSB/AIAA Student Panel, Washington, D.C.

Ocampo R. P. (2014, July). *The Limitations of Spacecraft Redundancy*. Presented at 44th International Conference on Environmental Systems (ICES), Tucson, AZ.

Ocampo R. P. (2015, September). *Linking Spacecraft Hazard Controls with System Design Requirements: General Considerations and Complications*. Presented at AIAA Space 2015 Conference and Exposition, Pasadena, CA.

Ocampo R. P. (2015, September). *The Space Shuttle's Commercial Potential: A Retrospective Analysis*. Presented at AIAA Space 2015 Conference and Exposition, Pasadena, CA.

Informal Presentations

Ocampo R. P. (2012, August). *Perspectives on Spacecraft Safety*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo R. P. (2012, December). *History of Spacecraft Safety: U.S. Manned Space Program*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo R. P. (2013, February). *Safety in Numbers: Evaluating Safety from a Quantitative Perspective*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo R. P. (2013, March). *History of Spacecraft Safety: Soviet/Russian Manned Space Program*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo R. P. (2013, October). *Spacecraft Redundancy*. Presented at Bioastronautics Student Seminar, Boulder CO.

Ocampo R. P. (2013, October). *Risk and Safety: Refining NASA's Nomenclature*. Presented at Bioastronautics Student Seminar, Boulder CO.

Conference Posters

Ocampo, R. P., & Klaus, D. M. (2012, July). *Defining a Safety Index for Human Spacecraft Design*. Poster session presented at the AIAA 42nd International Conference on Environmental Systems (ICES), San Diego, CA.

Ocampo, R. P., & Klaus, D. M. (2015, October). *Human Space Flight Safety*. Poster session presented at the FAA COE CST 5th Annual Technical Meeting, Washington, D.C.

REFERENCES

Aerospace Safety Advisory Panel. (1978). Aerospace Safety Advisory Panel Annual Report for 1977. *Washington, DC*.

Aerospace Safety Advisory Panel. (1979). Aerospace Safety Advisory Panel Annual Report for 1978. *Washington, DC*.

Aerospace Safety Advisory Panel. (1981). Aerospace Safety Advisory Panel Annual Report for 1980. *Washington, DC*.

Aerospace Safety Advisory Panel. (2002). Aerospace Safety Advisory Panel Annual Report for 2001. *Washington, DC*.

Aerospace Safety Advisory Panel. (2007). Aerospace Safety Advisory Panel Annual Report for 2006. *Washington, DC*.

Aerospace Safety Advisory Panel. (2009). Aerospace Safety Advisory Panel Annual Report for 2008. *Washington, DC*.

Aerospace Safety Advisory Panel. (2010). Aerospace Safety Advisory Panel Annual Report for 2009. *Washington, DC*.

Aerospace Safety Advisory Panel. (2011). Aerospace Safety Advisory Panel Annual Report for 2010. *Washington, DC*.

Aerospace Safety Advisory Panel. (2012). Aerospace Safety Advisory Panel Annual Report for 2011. *Washington, DC*.

Aerospace Safety Advisory Panel. (2014). Aerospace Safety Advisory Panel Annual Report for 2013. *Washington, DC*.

Aerospace Safety Advisory Panel. (2015). Aerospace Safety Advisory Panel Annual Report for 2014. *Washington, DC*.

Aerospace Safety Advisory Panel. (2016). Aerospace Safety Advisory Panel Annual Report for 2015. *Washington, DC*.

Amtrak. (2009). Amtrak System Timetables Fall 2009 Winter 2010.

Amtrak. (2010). Amtrak System Timetables Spring 2010 Summer 2010.

Amtrak. (2010). Amtrak System Timetables Fall 2010 Winter 2011.

Amtrak. (2011). Amtrak System Timetables Spring 2011 Summer 2011.

Amtrak. (2011). Amtrak System Timetables Fall 2011 Winter 2012.

Amtrak. (2012). Amtrak System Timetables Spring 2012 Summer 2012.

Amtrak. (2013). Amtrak System Timetables Winter 2013 Spring 2013.

Amtrak. (2013). Amtrak System Timetables Summer 2013 Fall 2013.

Amtrak. (2014). Amtrak System Timetables Winter 2014 Spring 2014.

Amtrak. (2014). Amtrak System Timetables Summer 2014 Fall 2013.

Ansoff, H. I. (1968). *Corporate strategy: Analytical approach to business policy*. Penguin.

Apollo 13 Review Board. (1970). Report of Apollo 13 Review Board. NASA.

Apollo 204 Review Board. (1967). Report of Apollo 204 Review Board to the Administrator National Aeronautics and Space Administration. Washington, DC: U.S. Government Printing Office.

Augustine, N., Austin, C. D. W., Bejmuk, M. B., Chyba, C., Crawley, E., Greason, M. J., & Kennel, C. (2009). Review of US Human Space Flight Plans Committee. *Seeking a Human Spaceflight Program Worthy of a Great Nation," NASA, Washington, DC*.

Belew, L. F., & Stuhlinger, E. (1973). Skylab: a guidebook.

Bergin C. (2011). Atlantis into down processing after MER review notes flawless return. *NASASpaceflight.com*. Retrieved from http://www.nasaspaceflight.com/2011/07/atlantis-down-processing-mer-review-notes-flawless-return/

Bilstein, R. E. (1980). Stages to Saturn: A Technological History of the Apollo/Saturn Launch Vehicles. NASA SP-4206. *NASA Special Publication*, *4206*.

Bogumil, R. J. (1982). Limitations of probabilistic risk assessment. *Technology and Society Magazine, IEEE*, *1*(3), 24-28.

Bond, A. (1988). A Review of Man-Rating in Past and Current Manned Space Flight Programs. *Houston: Eagle Engineering/LEMSCO*.

Bond, P. (2002). *The continuing story of the International Space Station*. Springer Science & Business Media.

Brooks, C. G., Grimwood, J. M., & Swenson Jr, L. S. (1979). Chariots for Apollo: a history of manned lunar spacecraft.

Bureau of Transportation Statistics. (2005). Issue Brief Number 13 - Airline Travel Since 9/11.

Burkhalter, B. & Sharpe, M. (1990). Mercury-Redstone: The first American man-rated space launch vehicle. Acta Astronautica, 21: 819-853.

Bureau of Transportation Statistics. (2016a). *Table 1-40: U.S. Passenger-Miles (Millions)* [Data file]. Retrieved from http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_01_40.html.

Bureau of Transportation Statistics. (2016b). *Table 2-9: U.S. Air Carrier(a) Safety Data* [Data file]. Retrieved from http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_02_09.html.

Bureau of Transportation Statistics. (2016c). *Table 2-14: U.S. General Aviation(a) Safety Data* [Data file]. Retrieved from http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_02_14.html

Bureau of Transportation Statistics. (2016d). *Table 2-17: Motor Vehicle Safety Data* [Data file]. Retrieved from http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_02_17.html.

Butler, R. W. (2008). A primer on architectural level fault tolerance.

Cassidy, J., Johnson, R., Leveye, J., & Miller, F. (1964). The Mercury-Redstone Project. *NASA*.

Chertok, B. E. (2009). *Rockets and People, Volume III, Hot Days of the Cold War*. Government Printing Office.

Columbia Accident Investigation Board. (2003). Report of the Columbia Accident Investigation Board. *Volume I*, 178.

Day, D. (2011). The Decision to Retire the Space Shuttle. *The Space Review*. Retrieved from http://www.thespacereview.com/article/1887/1

Dezfuli H. (2010). NASA's Risk Management Approach. Workshop on Risk Assessment and Safety Decision Making Under Uncertainty. Bethesda, MD.

Dezfuli, H., Benjamin, A., Everett, C., Smith, C., Stamatelatos, M., & Youngblood, R. (2011). NASA System Safety Handbook. Volume 1; System Safety Framework and Concepts for Implementation.

Denoble, P. J., Marroni, A., & Vann, R. D. (2011). Annual fatality rates and associated risk factors for recreational scuba diving.

Downer, J. (2009). *When Failure is an Option: Redundancy, reliability and regulation in complex technical systems*. Centre for Analysis of Risk and Regulation, London School of Economics and Political Science.

Embrey, S. (1966). The Apollo Saturn emergency detection system.

FAA. (2000). FAA System Safety Handbook.

FAA. (2004). *General Aviation and Part 135 Activity Surveys, CY 2003* [Data file]. Retrieved from https://www.faa.gov/data_research/aviation_data_statistics/general_aviation/CY2003/

FAA. (2005). *General Aviation and Part 135 Activity Surveys, CY 2004* [Data file]. Retrieved from https://www.faa.gov/data_research/aviation_data_statistics/general_aviation/CY2004/

FAA. (2006). *General Aviation and Part 135 Activity Surveys, CY 2005* [Data file]. Retrieved from https://www.faa.gov/data_research/aviation_data_statistics/general_aviation/CY2005/

FAA. (2007). *General Aviation and Part 135 Activity Surveys, CY 2006* [Data file]. Retrieved from https://www.faa.gov/data_research/aviation_data_statistics/general_aviation/CY2006/

FAA. (2008). *General Aviation and Part 135 Activity Surveys, CY 2007* [Data file]. Retrieved from https://www.faa.gov/data_research/aviation_data_statistics/general_aviation/CY2007/

Federal Railroad Administration Office of Safety Analysis. (2016). *Operational Data* [Data file]. Retrieved from http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx

Federal Railroad Administration Office of Safety Analysis. (2016). *Railroad Casualties* [Data file]. Retrieved from http://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx

Feynman, R. P. (1986). Personal observations on the reliability of the shuttle. *Report of the Presidential Commission on the Space Shuttle Challenger Accident*, *2*, 1-5.

Fragola, J. R. (1996, May). Risk management in US manned spacecraft: From Apollo to Alpha and beyond. In *Product Assurance Symposium and Software Product Assurance Workshop* (Vol. 377, p. 83).

Franzini, B. J., & Fragola, J. R. (2011, January). Human rating of launch vehicles: Historical and potential future risk. In *Reliability and Maintainability Symposium (RAMS), 2011 Proceedings-Annual* (pp. 1-6). IEEE.

French, J., & Bailey, F. Jr. (1963) Mercury Project Summary: Including Results of the Fourth Manned Orbital Space Flight. NASA.

Gibbons, J. (2008). Salyut: Soviet steps toward permanent human presence in space. DIANE Publishing.

Hacker, B. C., & Grimwood, J. M. (1977). On the Shoulders of Titans: A History of Project Gemini. *NASA, Rept. SP-4203*, 265-298.

Hall, R., & Shayler, D. J. (2001). *The Rocket Men: Vostok & Voskhod. The First Soviet Manned Spaceflights*. Springer Science & Business Media.

Hall, R., & Shayler, D. (2003). *Soyuz: a universal spacecraft*. Springer Science & Business Media.

Hamlin, T. L., Thigpen, E., Kahn, J., & Lo, Y. (2011, January). Shuttle Risk Progression: Use of the Shuttle Probabilistic Risk Assessment (PRA) to Show Reliability Growth. In *AIAA SPACE 2011 Conference & Exposition* (p. 7353).

Harland, D. M. (2007). *The story of space station MIR*. Springer Science & Business Media.

Harris, E. & Brom, J. (1965). Apollo launch-vehicle man-rating: Some considerations and an alternative contingency plan. NASA.

Heppenheimer, T. A. (2002). *The Space Shuttle Decision, 1965-1972*. Smithsonian Institution Press.

Hitt, D., Garriott, O. K., & Kerwin, J. (2008). *Homesteading Space: The Skylab Story*. U of Nebraska Press.

International Space Station Independent Safety Task Force. (2007). Final Report of the International Space Station Independent Safety Task Force.

Ivanovich, G. S. (2008). *Salyut-The First Space Station: Triumph and Tragedy*. Springer Science & Business Media.

Klaus, D., Ocampo, R., & Fanchiang, C. (2014, March). Spacecraft human-rating: Historical overview and implementation considerations. In *Aerospace Conference, 2014 IEEE* (pp. 1-7). IEEE.

Linenger, J. M. (2000). *Off the planet: Surviving five perilous months aboard the space station Mir*. McGraw Hill Professional.

List of rail accidents (2010-present). (n.d.). In *Wikipedia*. Retrieved November 9, 2015, from https://en.wikipedia.org/wiki/List_of_rail_accidents_(2010–present)

List of Russian manned space missions. (n.d.). In *Wikipedia*. Retrieved November 9, 2015, from https://en.wikipedia.org/wiki/List_of_Russian_manned_space_missions

List of Soviet manned space missions. (n.d.). In *Wikipedia*. Retrieved November 9, 2015, from https://en.wikipedia.org/wiki/List_of_Soviet_manned_space_missions

List of Space Shuttle missions. (n.d.). In *Wikipedia*. Retrieved November 9, 2015, from https://en.wikipedia.org/wiki/List_of_Space_Shuttle_missions

Low, G. M. (1970). What made Apollo a success? *Astronautics and Aeronautics*, *8*, 36-45.

McIntosh, S. E., Campbell, A. D., Dow, J., & Grissom, C. K. (2008). Mountaineering fatalities on Denali. *High Altitude Medicine & Biology*, *9*(1), 89-95.

Mikulak, R. J., McDermott, R., & Beauregard, M. (2008). *The basics of FMEA*. CRC Press.

Mulville, D. R. (1996). *Structural design and test factors of safety for spaceflight hardware*. Technical Report NASA-STD-5001, NASA.

Murray, C. A., & Cox, C. B. (1989). *Apollo, the Race to the Moon*. Simon & Schuster.

Musgrave, G. E., Larsen, A., & Sgobba, T. (2009). *Safety design for space systems*. Butterworth-Heinemann.

NASA. (1988). Guidelines for Man Rating Space Systems.

NASA. (2008a). NASA NPR 8705.2B. Human-Rating Requirements for Space Systems (w/Change 4 dated 8/21/12). *Washington, D.C*.

NASA. (2008b). NASA NPR 8715.3 NASA General Safety Program Requirements (w/Change 9 dated 2/08/13). *Washington, D.C*.

NASA. (2011). NASA/SP-2011-3422 NASA Risk Management Handbook. *Washington DC*.

NASA. (2015a). NASA CCT-REQ-1130 Rev D-1, ISS Crew Transportation and Services Requirements Document. *Washington, D.C*.

NASA. (2015b). NASA/SSP 50808 Rev E, International Space Station (ISS) to Commercial Orbital Transportation Services (COTS) Interface Requirements Document (IRD). *Washington, D.C*.

National Association of Railroad Passengers. (2015). Amtrak fact sheet, 2008-2014 [Data file]. Retrieved from https://www.narprail.org/our-issues/ridership-statistics/

National Highway Traffic Safety Administration. (2016). *Fatality Analysis Reporting System (FARS) Encyclopedia* [Data file]. Retrieved from http://www-fars.nhtsa.dot.gov/Main/index.aspx

National Park Service. (2010). Mount Rainier Annual Climbing Statistics.

National Parks Service (2011). Denali National Park and Preserve Annual Mountaineering Summary - 2010. *Washington, D.C*.

National Parks Service. (2012). Denali National Park and Preserve Annual Mountaineering Summary - 2011.

National Parks Service. (2013). Denali National Park and Preserve Annual Mountaineering Summary - 2012.

National Parks Service. (2014). Denali National Park and Preserve Annual Mountaineering Summary - 2013.

National Parks Service. (2015). Denali National Park and Preserve Annual Mountaineering Summary - 2014.

National Park Service. (2016). *Fatalities at Mt. Rainier National Park* [Data file]. Retrieved from http://www.mountrainierclimbing.us/sar/fatalities.php

National Transportation Safety Board. (2016). *Summary of US Civil Aviation Accidents for Calendar Year 2012* [Data file]. Retrieved from http://www.ntsb.gov/investigations/data/Pages/2012%20Aviation%20Accidents%20Summary.aspx

Ocampo, R. P. (2014). Limitations of Spacecraft Redundancy: A Case Study Analysis. 44th International Conference on Environmental Systems.

Ocampo, R. P. (2015). The Space Shuttle's Commercial Potential: A Retrospective Analysis. In *AIAA SPACE 2015 Conference and Exposition* (p. 4614).

Ocampo, R. P., & Klaus, D. M. (2016a). A Quantitative Framework for Defining "How Safe is Safe Enough" in Crewed Spacecraft. *New Space. Manuscript in press*.

Ocampo, R. P., & Klaus, D. M. (2016b). *Comparing the Relative Risk of Space Flight to Terrestrial Modes of Transportation and Adventure Sport Activities*. Manuscript in preparation.

Presidential Commission On Space Shuttle Challenger, & Rogers, W. P. (1986). Report of the presidential commission on the space shuttle challenger accident.

Rominger, K. V. (2004). Astronaut Office Position on Future Launch System Safety. *NASA JSC position paper submitted under cover memo CB-04044 from CB/Chief, Astronaut Office, to CA/Director, Flight Crew Operations*, *4*.

Salisbury, R., & Hawley, E. (2011). *The Himalaya by the numbers: a statistical analysis of mountaineering in the Nepal Himalaya*. Vajra Publications.

Santos, A., McGuckin, N., Nakamoto, H. Y., Gray, D., & Liss, S. (2011). *Summary of travel trends: 2009 national household travel survey* (No. FHWA-PL-ll-022).

Shelton, W. R. (1968). *Soviet Space Exploration: The First Decade*. Washington Square Press.

Slay, A. (1988). *Post-Challenger Evaluation of Space Shuttle Risk Assessment and Management*. National Academies.

Stamatelatos M. (2010). Safety Goals at NASA or How Safe is Safe Enough and How to Get There. Trilateral Safety and Mission Assurance Conference. Washington, D.C.

Stockton, W., & Wilford, J. N. (1981). *Spaceliner*. Times Books.

Swenson Jr, L. S., Grimwood, J. M., & Alexander, C. C. (1966). This New Ocean: A History of Project Mercury. NASA SP-4201. *NASA Special Publication*, *4201*.

Theurer, B. (1976). *Space Shuttle: A Case Study in Economic Analysis*. Defense Systems Management School, Fort Belvoir, VA.

Tumer, I., Barrientos, F., & Mehr, A. F. (2005, January). Towards risk based design (RBD) of space exploration missions: a review of RBD practice and research trends at NASA. In *ASME 2005 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference* (pp. 687-695). American Society of Mechanical Engineers.

United States Parachute Association. (2015). Skydiving Safety. Retrieved from http://www.uspa.org/facts-faqs/safety

United States Coast Guard. (2013). National Recreational Boating Survey 2012.

United States Coast Guard. (2013). Recreational Boating Statistics 2012.

U.S. House of Representatives (1986). Investigation of the Challenger Accident. *Washington DC*.

Vaughan, D. (1997). *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. University of Chicago Press.

Waterman, J. (1991). *Surviving Denali: A Study of Accidents on Mount McKinley, 1903-1990*. The Mountaineers Books.

Williamson, R. A. (1999). Developing the Space Shuttle.

Young, A. (2007). *Lunar and planetary rovers: the wheels of Apollo and the quest for Mars*. Springer Science & Business Media.

Zak, A. (2009). Russia to save its ISS modules. *BBC News*. Retrieved from http://news.bbc.co.uk/2/hi/science/nature/8064060.stm

Zupp, G. (1995). A perspective on the Human-Rating process of US spacecraft: Both past and present.

APPENDIX A: DESCRIPTION OF SELECT RISK REDUCTION TECHNIQUES

"We can lick gravity, but sometimes the paperwork is overwhelming."

-Wernher Von Braun

A.1 Redundancy

A redundant system is one that can achieve its intended function through multiple independent pathways or elements (FAA, 2000; NASA, 2008b). In crewed spacecraft, redundancy is typically applied to systems that are critical for safety and/or mission success[3,4]. Since no piece of hardware can be made perfectly reliable, redundancy—in theory—allows for the benign (e.g., non-catastrophic) failure of critical elements.

Redundant elements can be 1) similar or dissimilar to each other, 2) activated automatically ("hot spare") or manually ("cold spare"), and 3) located together or separated geographically (Butler, 2008; Downer, 2009; Low, 1970). U.S. spacecraft have employed redundancy on virtually all levels of spacecraft design, from component to subsystem (Low, 1970; Hitt et al., 2008).

Redundancy has a successful history of precluding critical and catastrophic failures during human spaceflight. A review of NASA mission reports, from Mercury to Space Shuttle, indicates that redundancy has saved the crew or extended the mission over 160 times, or roughly once per flight. The presence of a partially redundant spacecraft (the Lunar Module) during the Apollo 13 emergency notably contributed to the crew's safe return (Apollo 13 Review Board, 1970).

A.2 Failure Tolerance

A failure tolerant system is one that can continue operating in the presence of one or more failures. In this regard, a redundant system is failure tolerant to a specific hazard. However, the reverse is not necessarily true, as failure tolerance can be achieved using techniques other than redundancy. For example, rather than add redundant thrusters to the spacecraft, consumables on board Vostok and Voskhod were sized to last until the natural decay of the vehicle's orbit. This allowed the spacecraft to be failure tolerant to loss of thruster capability.

A.3 Design for Minimum Risk (DFMR)

Design for Minimum Risk (DFMR) is a technique used to reduce hazards through increased margins. In cases where DFMR is solely employed, redundancy is impractical and/or counterproductive, (e.g., the redundant component increases complexity and/or consumes additional resources). Instead, strengthening (factor of safety), buffering (design margins), or more thorough evaluation (reliability analysis) of the DFMR component is applied.

A.4 Factors of Safety

Crewed spacecraft typically add a "factor of safety" to critical components, particularly those whose design does not allow for redundancy (e.g., tanks, windows). A component with a 2:1 factor of safety, for example, can theoretically withstand twice its anticipated loads.

This factor of safety allows the vehicle to survive unanticipated environmental conditions, uncertainty in design, and/or degradation over time, but comes at the cost of added

weight, expense, and complexity. Generally speaking, the higher the factor of safety, the safer the system[12].

The factor of safety chosen for a given component is dependent on the component's materials, attachment methods, and verification approach (Mulville, 1996). NASA typically employs a minimum 1.4:1 factor of safety on all flight hardware, regardless of whether the vehicle is manned or unmanned (Mulville, 1996). In areas where weight is less critical, higher factors of safety are employed. Spacecraft ground systems, for example, utilize a 4:1 factor of safety.

A.5 Quality Assurance

Quality assurance (QA) programs (e.g., testing, evaluation, and inspection) have been shown to significantly improve vehicle reliability. In those instances where testing and thorough evaluation of inspection or test results has been limited or curtailed, (e.g., Mercury, Vostok/Voskhod, pre-*Challenger* Space Shuttle, early Soyuz), risk has increased significantly, in some cases to the point of catastrophe (Swenson et al., 1966; Rogers et al., 1986; Chertok, 2009; Vaughan, 1996).

---

[12] There are certain, very limited, exceptions to this heuristic. Consider two pressurized vessels, one a steel tank and one a balloon. If the steel tank were to rupture, the resulting effects would be more harmful than if the balloon were to pop. However, this example is relevant only *after* a failure has occurred; as such, this heuristic is generally robust to most estimates of risk.

A.6 Operational Procedures

When design techniques fail to eliminate a hazard, operational procedures can be used to save crewmembers, particularly during long duration missions. Skylab, Salyut, Mir and ISS missions have all been saved or extended because of in-flight uploaded software patches or operational workarounds.

A.7 Ejection/Abort/Mission Termination

Ejection seats and/or abort modes serve as "last resorts" that may be used to prevent loss of crew in the case of catastrophic vehicle destruction. Although aborts have only been performed three times in the last half century (Soyuz T-10a pad abort; Soyuz 18a; STS-51-F Abort To Orbit), they have saved the crew on each occasion. Missions can also be prematurely terminated to ensure the crew's safe return.

APPENDIX B: DESCRIPTION OF SELECT SPACECRAFT SAFETY ANALYSES

B.1 Hazard Analysis (HA)

A Hazard Analysis (HA) is a systematic, top-down approach to identifying events or conditions (e.g., hazards) that can trigger undesirable outcomes. These hazards can be products of the flight hardware or threats posed by the environment or mission (Slay, 1988). Once these hazards are identified, they are qualitatively evaluated in terms of both their severity and their likelihood; hazards that are severe (e.g., hazards that can lead to Loss of Mission, Loss of Vehicle, or Loss of Crew) and/or highly likely are typically given the most scrutiny. Controls that eliminate or mitigate the risk are then identified; if no controls are available, the system is re-designed or the risk is accepted (FAA, 2000).

Hazard analyses have been performed on all U.S. human space programs, from Mercury to Constellation. However, because hazard analyses are qualitative in nature, they are frequently used in conjunction with other *quantitative* analyses, such as Probabilistic Risk Assessment (PRA).

B.2 Failure Modes and Effects Analysis (FMEA)

In a Failure Modes and Effects Analysis (FMEA), potential failure modes and their effects on the larger system are identified, and the *severity*, *occurrence* and *likelihood* of detection are evaluated on a 1-10 scale (Mikulak, 2008). These three values are then multiplied to produce a Risk Priority Number (RPN). Failure modes with relatively high RPNs (or with large category values), are generally assigned corrective action so as to reduce the overall RPN value (and/or the category value).

In instances where the RPN value cannot be significantly reduced, the potential failure mode can be recorded on a "Critical Items List" (CIL); its associated component is then monitored throughout the space flight or program (Fragola, 1996).

Although FMEA has a quantitative component to it, it is considered a qualitative analytical technique, as there is no set numerical cutoff to identify "risky" failure modes from "safe" failure modes (Fragola, 1996). Additionally, FMEA has trouble modeling dynamic situations, and cannot deal well with uncertainties in categorical value estimates.

B.3 Fault Tree Analysis (FTA)

A Fault Tree Analysis (FTA) is a top-down approach to identifying root cause(s) of undesirable events. In an FTA, an undesirable event—for example, Loss of Mission (LOM), Loss of Vehicle (LOV), or Loss of Crew (LOC)—is first designated the top 'shoot'; intermediate events and 'root' causes (e.g., basic events) are then identified via a backward-stepping process. Graphically distinct 'gates'—Or, And, Exclusive Or, Priority And, and Inhibit—are used to symbolize the boolean relationship between higher level-events and lower-level causes (Vesley et al., 2002)

Fault trees can be used in conjunction with a number of other analytical techniques. For example, top events identified in a Failure Modes and Effects Analysis (FMEA) can be analyzed with an FTA to identify root causes; these root causes can in turn be mapped to a corresponding Hazard Analysis (HA). Although fault trees are generally considered to be a qualitative analytic technique, they can also be evaluated *quantitatively* when probabilistic data are available, as described in the section below.

Generally, top events are engendered by relatively few root causes (Vesley et al., 2002). By using FTA to identify the *most problematic causes*, minimal resources can be used to mitigate most key safety issues.

B.4 Probabilistic Risk Assessment (PRA)

To calculate a PRA value, Initiating Events (IE) that can potentially lead to a catastrophic accident are identified using a Master Logic Diagram (MLD). Each initiating event is then mapped out to the catastrophic event (potentially via intermediate *pivotal events*) using an Event Sequence Diagram (ESD). Finally, probabilistic values are assigned to each node in the ESD, and the overall probability of the catastrophic event occurring due to the initiating event is calculated. Each IE is then summed to identify the overall probability of catastrophe.

While Probabilistic Risk Assessment (PRA) was available to engineers during the lead-up to the Apollo lunar landing, this technique were for the most part ignored, as the values associated with the analysis were considered to be too "pessimistic" (Fragola, 1996). At the time of the lunar landing, PRA identified the probability of mission success for the Saturn V rocket, the Command and Service Module (CSM), and Lunar Module (LM) to be 0.88, 0.90, and 0.95, respectively, for an overall mission success PRA of 0.75 (0.88 x 0.90 x 0.95). Because this calculated probability of mission success was significantly lower than the actuarial distribution of mission success (0.75 vs. 0.875), PRA techniques were abandoned until after the Challenger accident.

In large part due to the efforts of physicist Richard Feynman (Fragola, 1996, Feynman 1986), PRA became a prominent component of spacecraft safety analysis post-Challenger.

Unlike HA, FMEA, or FTA, Probabilistic Risk Assessment is considered a quantitative analysis; there are no subjective cutoffs (*a la* RPN), only analytically-calculated probability distributions.

APPENDIX C: DESCRIPTION OF SELECT RISK TERMINOLOGY

Fatality: A death that *directly* results from performing or engaging in an activity[13].

Fatal Accident: An accident in which one or more fatalities occur. No qualification is given to the total *number* of fatalities that occur during the accident.

Vehicle-Trip: A single, uninterrupted trip on board a single vehicle, regardless of exposure length or number of people on board. For example, a scheduled train trip from New York to Washington, D.C. would be considered a single vehicle-trip, even if the number of passengers change from station to station.

Vehicle-Hour: A single vehicle-hour is the movement of one vehicle for one hour, independent of the number of people on board vehicle. If *Bus A* is driven 3 hours and *Bus B* is driven 4 hours, the total number of vehicle-hours for the two buses is 7 hours.

Vehicle-Mile: A single vehicle-mile is the movement of one vehicle one mile, regardless of the number of people on board (Santos et al., 2011). A train that travels 100 miles on Saturday and 100 miles on Sunday has traveled a total of 200 vehicle-miles during the two-day period.

---

[13] The time period in which death must occur to be considered an activity-related fatality (e.g., within *x* days of the activity) is not specified here, as fatality data sources do not consistently define this time period. However, it may broadly be interpreted as occurring within 30 days of the activity, as this is the time frame specified by both the National Highway Traffic Safety Association (NHTSA) and the Federal Aviation Administration (FAA) (NHTSA, FAA)

Person-Trip: A single person-trip is the exposure of one person to one trip (or activity). An individual that dives the same reef 4 times accounts for 4 person-trips (1 person x 4 trips = 4 person-trips); 3 people who take 2 sailing trips result in 6 person-trips (3 persons x 2 trips = 6 person-trips)[14].

Person-Hour: A single person-hour is the movement of one person for one hour. An aircraft that flies 200 people for 5 hours results in 1,000 person-hours (200 people x 5 hours).

Person-Mile: A single person-mile is the movement of one person one mile. A 10-mile car trip with 3 passengers on board would accumulate 30 person-miles (3 people x 10 miles).

---

[14] With transportation data, person-trips are equivalent to the number of non-unique passengers that have participated in an activity during a given time period; however, the term person-trip is retained here to account for the fact that participants in adventure sport activities are generally not referred to as "passengers."

"But he supposed that even as bad as it was then, there must be some kind of pleasure mixed in with it, some desperate relief. Joy, perhaps, in knowing that it was over, one more time, with nothing held back."

- Quenton Cassidy