A Quantitative Framework for Defining "How Safe is Safe Enough?" in Crewed Spacecraft

Robert P. Ocampo and David M. Klaus

Aerospace Engineering Sciences, University of Colorado Boulder, Boulder, Colorado.

ABSTRACT

*This article presents a quantitative framework that can be used to determine whether a spacecraft is "Safe Enough" for crewed flight. Three major elements are established, namely: (1) An unequivocal definition of "Safe" (and its inverse, "Unsafe"), derived from and consistent with current NASA terminology and empirical practice, (2) a quantitative risk spectrum, which highlights spacecraft risk (measured using Probabilistic Risk Assessment) against a reference standard, and (3) a probabilistic threshold value, which delineates risk that is acceptable (*e.g.*, "Safe Enough") with risk that is unacceptable (*e.g.*, "Not Safe Enough"). Each element of the framework is developed concomitantly, step-wise, and from the bottom-up, to ensure "Safe Enough" can be reliably and systematically determined.*

> *"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?"*[1]
>
> —Glynn Lunney, Apollo Flight Director

INTRODUCTION

Spaceflight is an inherently risky endeavor. Recent accidents, including the loss of Orbital Sciences' uncrewed Cygnus spacecraft, the destruction of Space X's uncrewed Dragon vehicle, and the in-flight death of a Virgin Galactic test pilot serve to underscore this point. But how risky is *too* risky? Conversely, how safe is safe *enough*?

These questions are not new ones. As early as Project Mercury, space officials were asking "How simple is safe?"[2] and "What…does 'good' mean?"[1] However, in the past decade these questions have grown in urgency: The Aerospace Safety Advisory Panel (ASAP)—the independent advisory panel charged by Congress with evaluating NASA's safety performance—has posed some form of the question "How safe is safe enough" in *six of its last seven annual reports.*[3–8]

To answer this question, a means of distinguishing "Safe Enough" from "Not Safe Enough" is required. For much of its history, NASA has relied on hazard analyses to delineate these divergent system states.[9] Under this rubric, a spacecraft is deemed to be "safe"* if all hazards (*e.g.*, conditions that can trigger an undesirable outcome) have a corresponding hazard control in place to eliminate or mitigate the hazard's likelihood or severity.[9]

Although this hazards-based rubric is capable of identifying hazards associated with *individual* components (or subsystems), it is not well suited for evaluating hazards that arise from *systemic* failures.[9,10] For this reason, NASA has shifted to a more holistic, requirements-based methodology for evaluating risk. 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.[11,12]

A proscriptive, requirements-based 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.[11,13,14] 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, the use of FT to protect against one hazard can actually increase the likelihood of another hazard occurring—which, in turn, can lead to an overall *increase* in system risk. For example, the addition of a redundant depressurization valve on Soyuz 11 was intended to protect the crew against pressure equalization failures (*i.e.*, a valve "failed closed" event) during reentry and landing, but ultimately contributed to the vehicle's catastrophic depressurization and loss of crew (LOC) when it failed *open* in flight, allowing the cabin atmosphere to vent overboard.[15] Although it might be argued that this accident was due to inadequate system design or verification, it nonetheless provides an example where adding a component intended to provide FT to one hazard increased the likelihood of a second hazard occurring.

*In this context, "safe" can be considered analogous to "Safe Enough."

DOI: 10.1089/space.2015.0040

Under circumstances 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 *less safe* 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" methodologies. Given the limitations inherent to both hazard-based and requirements-based "Safe Enough" methodologies, this article proposes an alternative framework for discriminating "Safe Enough" spacecraft from "Not Safe Enough" spacecraft. Three major elements are established, namely:

1. Unambiguous *definitions* of "Safe," "Unsafe," and "Risk." (The *Safe* part of "How Safe is Safe Enough?")
2. A consistent, quantitative *risk metric* and corresponding *risk spectrum* for measuring and assessing risk. (The *How* part of "How Safe is Safe Enough?")
3. A specified *threshold*, located on the risk spectrum, which serves to delineate a "Safe Enough" spacecraft from a "Not Safe Enough" spacecraft. (The *Enough* part of "How Safe is Safe Enough?") This threshold considers "Safe Enough" in terms of what is both *acceptable* and *achievable*. An *acceptable* level of risk is determined by a programmatic or personal decision, whereas an *achievable* level of risk is a function of engineering practice, available budget, schedule targets, cumulative operational experience, and other factors.

Although a similar framework for distinguishing spacecraft has been alluded to elsewhere,[16,17] to the authors' knowledge, the "Safe Enough" framework explicitly described 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?" The foundation of this framework starts with establishing unambiguous definitions of the terms "Safe," "Unsafe," and "Risk."

DEFINING "SAFE," "UNSAFE," AND "RISK"

Limitations of Current Definitions

In their 1978 annual report, ASAP stated that one of the primary obstacles to defining "Safe Enough" spacecraft stems from the ambiguous use of the term "safety" 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.*"[18]

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 before *Challenger*), the Space Shuttle was becoming "increasingly *unsafe*"[19] [emphasis added]. 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 that the vehicle was growing increasingly unsafe.[†] 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*"[20] [emphasis added].

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. To successfully determine whether a spacecraft is "Safe Enough," unambiguous definitions of "Safe" and "Unsafe" must first be established.

Developing New Definitions

NASA's current definition of "Safety" serves as an ideal starting point from which to build upon, as it serves to implicitly describe "Safe" and "Unsafe." According to NASA's General Safety Program Requirements,[21] "Safety" is:

> *"Freedom from those conditions that can cause death, injury, occupational illness, damage to or loss of equipment or property, or damage to the environment."*[21,‡]

Given this definition, a spacecraft can, therefore, either be:

Safe (NASA-derived definition): System 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."

[†]Although the final report of the Columbia Accident Investigation Board (CAIB) explicitly stated that the Shuttle was not "inherently unsafe,"[22] some CAIB members have gone on record to say that their recommendation to recertify the Space Shuttle in 2010 constituted an implicit concern over the vehicle's ability to fly safely beyond this timeframe.[23]

[‡]The full definition from NASA NPR 8715.3C adds that "In a risk-informed context, safety is an overall mission and program condition that provides sufficient assurance that accidents will not result from the mission execution or program implementation, or, if they occur, their consequences will be mitigated. This assurance is established by means of the satisfaction of a combination of deterministic criteria and risk criteria."[21] This addendum, however, is more in line with the concept of "Safe Enough," which is addressed (and whose definition is reached) later in this article.

or

Unsafe (NASA-derived definition): System 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."

This classification system allows spacecraft to be categorized with nearly perfect certainty: Since no real-world spacecraft can ever be completely free from "conditions" that can cause harm,[24,25,§] no spacecraft can ever be classified as "Safe." Therefore, *all* spacecraft must be considered "Unsafe" by default. Moreover, all spacecraft must be considered *uniformly* "Unsafe" by this classification system, as the discrete definition provided by NASA does not differentiate between varying *degrees* of "Unsafe."

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"[21]) is generally perceived 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 1 *unlikely* condition (*e.g.*, 1% likelihood).

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

Safe (Revised Definition 1): System 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.*, nontheoretical) system can ever be free of such "conditions," this state is unachievable.

Unsafe (Revised 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 (*e.g.*, conditions that can result in "fatal injury, loss of vehicle, or permanently disabling injury," as previously mentioned[21]) 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"[21]). This article focuses specifically on *catastrophic* conditions, as these conditions tend to be of primary concern to NASA and other space agencies**; as such, the definitions of "Safe" and "Unsafe" are further modified as follows for this framework:

Safe (Revised Definition 2): System is free from all catastrophic conditions. Given that no practical (*e.g.*, nontheoretical) system can ever be free of such "conditions," this state is unachievable.

Unsafe (Revised 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."

Concise (Working) Definitions

The terminology used to define "Safe" and "Unsafe" as mentioned is precise but unwieldy. To simplify these definitions, the term "hazard" will be used instead of "conditions…," because the two terms are virtually synonymous per NASA's definitions (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"[21]).

In addition, 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,"[21] "degrees of unsafe" can (and will) be articulated as "Risk" throughout this article. A final definition of "Safe," "Unsafe," and "Risk" is found hereunder; for a more detailed description of the evolution of the terms, see *Table 1*.

Safe (Revised Definition 3—Final Definition): System is free from all catastrophic hazards. Given that no practical (*e.g.*, nontheoretical) system can ever be free of such hazards, this state is unachievable.

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

**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.

Table 1. Evolution of "Safe" and "Unsafe" Definitions

Version	Reason for Update	Definition of Safe	Definition of Unsafe
Baseline (via NASA NPR 8715.3C)	N/A	System 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.[21]	System 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.[21]
Rev 1	Since no spacecraft can ever be free from conditions that can cause harm,[24,25] no spacecraft can ever be considered "Safe."	System 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 system can ever be free of such "conditions," this state is unachievable.*	NO CHANGE
Rev 2	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."*
Rev 3	The severity of the conditions must be specified to ensure that spacecraft are assessed against equivalent standards. The definition listed in this article specifies "catastrophic" conditions, as these are of primary concern to space agencies.	System 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."
Rev 4	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.	System 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."

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."[21] Therefore, "degrees of unsafe" will be articulated as "degrees of risk" throughout the remainder of this article; however, the final definition of "Unsafe" does not change. When new aspects of the definitions are added, they are depicted in italics.

Unsafe (Revised 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."

ASSESSING AND QUANTIFYING RISK

Visual Framework

The probability that a spacecraft will experience a catastrophic hazard (or set of hazards) naturally ranges from zero to unity (noninclusive). Spacecraft "Risk" can, therefore, be depicted as a spectrum of values per *Figure 1*, ranging from (but not including) 0% to 100% likelihood. The lower limit of this spectrum (*e.g.*, 0% risk) represents the unachievable "Safe" state. The remainder of the spectrum (*e.g.*, risk >0%) represents the "Unsafe" state.

Establishing a Quantitative Risk Metric

A spacecraft's position on the "Unsafe" segment (*e.g.*, its "Risk") can be estimated using a technique known as Probabilistic Risk Assessment (PRA). PRA predicts the likelihood that a hazard (or set of hazards) will occur by creating and assessing a mathematical logic model of a physical spacecraft.[26] Failure probabilities are determined for individual components and/or events, then amalgamated within the model to produce an overall estimate of mean (*e.g.*, average) risk and uncertainty.[27]

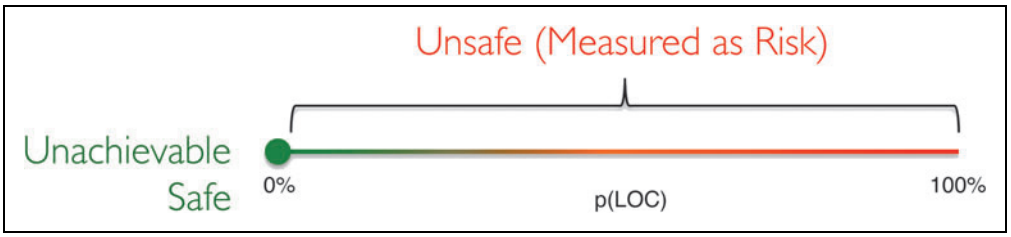

Fig. 1. Catastrophic risk (*e.g.*, the degree to which a spacecraft is "unsafe") can be characterized as a spectrum as in this figure. The probability of LOC—p(LOC)—serves to define where a spacecraft lies on this spectrum. Lower average p(LOC) values represent lower risk (*left side* of the spectrum), whereas higher average values represent higher risk (*right side* of the spectrum). An average p(LOC) of 0% is equivalent to the "Safe" state and is considered unachievable in this framework. LOC, loss of crew.

When PRA is used to quantify *catastrophic* risk, these two values (mean and uncertainty) are collectively referred to as probability of LOC, or p(LOC), values. A mean p(LOC) value near 0% is indicative of low spacecraft risk and appears on the far left side of the spectrum; conversely a mean p(LOC) value near 100% is indicative of high spacecraft risk and appears on the far right side of the spectrum[††] (*Fig. 1*).

Limitations of PRA

PRA is not a foolproof means of determining the true risk of a spacecraft, in part, because the quantitative failure rate data used to drive the analysis tend to be focused on design (as opposed to the process) failures and because such failure rate data may not be accurate or complete.[28–32] Catastrophic spaceflight failures are (fortunately) infrequent, but this can be a double-edged sword: the very infrequency makes accurately quantifying failure rates extremely difficult, particularly during the early stages of spacecraft development. Although some engineers claim that probabilistic methods can handle low failure rate data using Bayesian methods,[33,24] others contend that accurate PRA values cannot be determined without *empirically* verifying each component's life expectancy (which requires many hundreds of hours of testing). Because the majority of spacecraft have millions of parts that can fail in tens (or hundreds) of different ways, some have argued that it is better to focus resources on fixing design flaws than on probabilistically measuring them.[1,34]

Benefits of Using PRA to Evaluate Risk

Despite these limitations, PRA is currently the optimal method for evaluating a spacecraft's position on the risk spectrum. Although NASA relies on a combination of different methodologies to *qualitatively* evaluate risk—including hazard analysis, fault tree analysis, and failure modes and effects analysis—PRA is one of the few methodologies capable of building upon these qualitative techniques to effectively *quantify* risk.

Moreover, PRA estimates of risk tend to be more precise and accurate than risk estimates quantified using other methodologies, such as expert opinion. Before the *Challenger* accident (and NASA's initial use of PRA), the estimated likelihood of a catastrophic Space Shuttle accident ranged from 1 in 100 to 1 in 100,000—*a range of three orders of magnitude*.[35] Later estimates, which incorporated PRA techniques, placed the mean risk value between 1/60 and 1/78.[36] The latter two estimates constitute a much smaller range of values and, more importantly, better resemble the Space Shuttle's *actuarial*[‡‡] risk of 1/67.5 (*e.g.*, two catastrophic accidents in 135 flights).

In addition, PRA estimates of mean risk do not need to be perfect to have utility. PRA uncertainty measurements establish a range of values, roughly centered around the mean, that (likely) encapsulate the true risk of the spacecraft. If PRA uncertainty is low, the range of values that bound the true risk is small; if PRA uncertainty is high, the range of values that bound the true risk is large. In either case, PRA uncertainty (in conjunction with the PRA mean) can be readily used to establish a spacecraft's *range of* locations on the risk spectrum; this, in turn, is sufficient information to assess whether a spacecraft is "Safe Enough" for flight (see Specifying a "Safe Enough" Risk Threshold section).

SPECIFYING A "SAFE ENOUGH" RISK THRESHOLD

Given the characteristics of the risk spectrum described previously, the primary variable of interest—"Safe Enough"—can now be assessed. Spacecraft that exhibit mean p(LOC) values statistically less than or equal to an established risk threshold (which can be determined using the p(LOC) uncertainty value) can be considered "Safe Enough"; conversely, spacecraft that *do not* exhibit mean p(LOC) values statistically less than or equal to this threshold can be considered "Not Safe Enough" (*Fig. 2*).

[††]Mean p(LOC) values provide an indication of spacecraft risk but serve only as point estimates of the spacecraft's *true* risk. p(LOC) uncertainty measurements provide additional fidelity to the analysis and are discussed further in the Benefits of Using PRA to Evaluate Risk section.

[‡‡]PRA produces probabilistic values, which can differ (sometimes drastically) from measured actuarial values. However, a high-fidelity estimate of p(LOC) should ultimately converge (over time) with the spacecraft's actuarial rate of LOC.

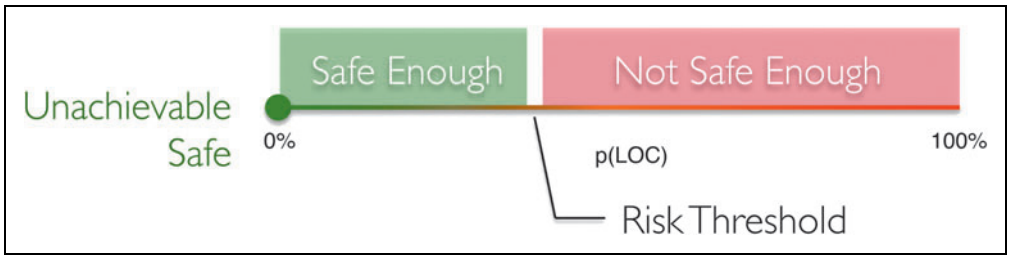

Fig. 2. Complete framework for distinguishing "Safe Enough" from "Not Safe Enough." Spacecraft with a p(LOC) less than or equal to the risk threshold (with a specified level of statistical certainty) can be considered "Safe Enough." Spacecraft with a p(LOC) that is NOT less than the risk threshold (with a specified level of statistical certainty) can be considered "Not Safe Enough."

Under this rubric, calculating whether a spacecraft is "Safe Enough" is mathematically simple. *Establishing an appropriate threshold value*, however, constitutes a far greater challenge. The chosen threshold value must balance what is desirable (or acceptable) from a programmatic or personal standpoint (*e.g.*, lower risk thresholds) with what is achievable given the program's budget, schedule, and engineering capabilities.[5] Selecting too low a threshold diverts resources toward a goal that may be unachievable; conversely, selecting too high a threshold places unnecessary risk on both the crew and the spacecraft.[§§]

NASA has established 1/200 as the overall target p(LOC) value for commercial spacecraft traveling to the International Space Station (ISS).[13] This value is five times riskier than the 1/1000 p(LOC) value that was originally required for Constellation's ISS mission[37,***]; however, it also represents a roughly twofold *reduction* in risk compared to the Space Shuttle, which exhibited an overall mean p(LOC) value of 1/90 toward the end of its career.[38]

Determining whether these (or any) threshold values are appropriate is beyond the scope of establishing this framework. Different programs, flying different missions of various durations, may be willing to accept more or less risk (and more or less uncertainty within the statistical assessment), which may also be distributed differently over each phase of flight. Regardless of *what* threshold value is chosen, however, the functionality of the framework (if not the results) remains the same.

[§§]Selecting an appropriate value to define statistical significance is similarly complex. If too small a significance level is chosen (*e.g.*, $P \leq 0.01$), "Safe Enough" may become difficult, if not impossible to achieve. Conversely, if too high a significance level is chosen (*e.g.*, $P \leq 0.1$), spacecraft considered statistically "Safe Enough" may not be "Safe Enough" in reality.

***To the authors' knowledge, Constellation was the first program to establish a target threshold p(LOC) value for its crewed missions.

RESULTS

This framework defines "Safe Enough" spacecraft as those that exhibit p(LOC) values statistically less than or equal to an established risk threshold. Consequently "Safe Enough" can be *mathematically* achieved in one of three ways:

1. *Reduce spacecraft mean p(LOC).* Reducing a spacecraft's mean p(LOC) to below its risk threshold can serve to shift the spacecraft's state from "Not Safe Enough" to "Safe Enough." For example, although there was never a specified "Safe Enough" threshold established for the Space Shuttle, the vehicle's measured mean p(LOC) was substantially improved over the course of its history, from 1/12 for STS-1 to 1/90 by STS-133.[38] This was accomplished by implementing a number of design and operational changes that reduced risk.
2. *Reduce uncertainty in the spacecraft risk analysis.* Although reducing the uncertainty of the risk analysis does nothing to change the *true* risk of a spacecraft, uncertainty reduction can alter the *estimate* of risk in a manner that changes the assessment of a spacecraft from "Not Safe Enough" to "Safe Enough" (or vice versa). As experience was gained over the course of the Space Shuttle program, risk uncertainty was reduced by roughly an order of magnitude,[38] allowing estimates of p(LOC) to become substantially more accurate.
3. *Increase acceptable risk thresholds.* By shifting the required risk threshold toward the right (*i.e.*, acceptance of greater risk), spacecraft with higher mean p(LOC) values (and potentially greater uncertainty) can be accepted as "Safe Enough." This technique was implemented in 2010, when NASA increased Constellation's ISS mission risk threshold from 1/1000 to 1/270,[37] and again in 2015, when the Commercial Crew Program increased its risk threshold from 1/270[39] to 1/200.[13]

It should be noted that the first two "Safe Enough" techniques previously listed will not necessarily be achievable by every program. Reducing a spacecraft's mean p(LOC) to below a set threshold may be impossible, given the inherent extremes of the spaceflight environment and the limits of present-day design and manufacturing techniques. Decreasing risk uncertainty to an acceptable level may also be infeasible because of budget and schedule constraints, which preclude extensive analysis and/or repeated testing of the spacecraft and its

subsystems. Nevertheless, these first two techniques—which *reduce* the true or estimated risk of the spacecraft—may be preferable to the third approach, which simply *accepts* greater risk for the program.[†††]

CONCLUSIONS

By definition, no spacecraft can ever be perfectly "Safe." Therefore, engineers must strive for "Safe Enough"—defined here as exhibiting a mean p(LOC) value less than or equal to a specified threshold (with a given level of statistical certainty). This can be achieved by (1) reducing the spacecraft's mean p(LOC), (2) reducing the spacecraft's p(LOC) uncertainty, and/or (3) increasing the acceptable risk threshold. As the United States works to develop the next generation of crewed spacecraft in both the government and commercial sectors, these three methods offer distinct—yet complementary—approaches to designing, constructing, flying, and maintaining "Safe Enough" spacecraft.

ACKNOWLEDGMENT

The FAA has sponsored this project, in part, through the Center of Excellence for Commercial Space Transportation. However, the agency neither endorses nor rejects the findings of this research. The presentation of this information is in the interest of invoking technical community comment on the results and conclusions of the research.

AUTHOR DISCLOSURE STATEMENT

No competing financial interests exist.

REFERENCES

1. Murray C, Cox C. Apollo, the Race to the Moon. New York: Simon & Schuster, 1989.
2. Swenson L, Grimwood, J, Alexander C. This New Ocean: A History of Project Mercury. Washington, DC: NASA, 1966.
3. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 2008. 2009.
4. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 2009. 2010.
5. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 2010. 2011.
6. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 2012. 2013.
7. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 2013. 2014.
8. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 2014. 2015.
9. NASA. NASA System Safety Handbook: Volume 1, System Safety Framework and Concepts for Implementation. NASA/SP-2010-580, 2011.
10. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 2006. 2007.
11. NASA. Human-Rating Requirements for Space Systems (w/change 4 dated 8/21/2012). NASA NPR 8705.2B, 2008.
12. Klaus D, Ocampo R, Fanchiang C. Spacecraft human-rating: Historical overview and implementation considerations. IEEE Aerospace Conference Proceedings. Big Sky, MT, 978-1-4799-1622-1/14, no. 2272, 2014.
13. NASA. ISS Crew Transportation and Services Requirements Document. NASA CCT-REQ-1130 Rev D-1, 2015.
14. NASA. International Space Station (ISS) to Commercial Orbital Transportation Services (COTS) Interface Requirements Document (IRD). NASA SSP 50808 Rev C, 2011.
15. Ocampo R. Limitations of spacecraft redundancy: A case study analysis. 44th International Conference on Environmental Systems. Tucson, AZ, ICES-2014-248, 2014.
16. Dezfuli H. NASA's risk management approach. Workshop on Risk Assessment and Safety Decision Making Under Uncertainty, Bethesda, MD, 2010.
17. Stamatelatos M. Safety goals at NASA or how safe is safe enough and how to get there. Trilateral Safety and Mission Assurance Conference, Washington, DC, 2010.
18. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 1978. 1979.
19. Committee on Science and Technology House of Representatives. Investigation of the Challenger Accident. Washington, DC: US Government Printing Office, 1986.
20. Bergin C. 2011. Atlantis into down processing after MER review notes flawless return. www.NASASpaceflight.com. (last accessed March 18, 2014).
21. NASA. NASA General Safety Program Requirements. NASA NPR 8715.3C, 2008.
22. Columbia Accident Investigation Board. CAIB Report. Washington DC: NASA, 2003.
23. Day D. 2011. The decision to retire the Space Shuttle. The Space Review. (last accessed March 18, 2014).
24. Committee on Shuttle Criticality Review and Hazard Analysis Audit. Post-Challenger Evaluation of Space Shuttle Risk Assessment and Management. Washington DC: National Academy Press, 1988.
25. Aerospace Safety Advisory Panel. Aerospace Safety Advisory Panel Annual Report for 1980. 1981.
26. Bogumil R. Limitations of probabilistic risk assessment. Technol Soc Mag. 1982;1: 24–28.
27. NASA. Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners. Washington, DC: NASA, 2002.
28. Freudenberg W. Heuristics, biases, and the not-so-general publics: Expertise and error in the assessment of risks. In: Social Theories of Risk. Westport, CT: Praeger 1992.
29. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. Science. 1988;185(4157):1124–1131.
30. Fischhoff B, Lichtenstein S, Slovic P, et al. Acceptable Risk. New York: Cambridge University Press, 1981.
31. Lowrance, W. Of acceptable risk: Science and the Determination of Safety. Los Altos, CA: William Kaufmann, 1976.
32. Mahler J. Organizational Learning at NASA: The Challenger and Columbia Accidents. Washington, DC: Georgetown University Press, 2009.
33. Rutledge P, Buchbinder B. A quantitative, probabilistic approach to human-rating space systems. Reliability and Maintainability Symposium Annual Proceedings, Anaheim, CA, 1994.

[†††]This is not to say that increasing a program's risk thresholds should be construed as cavalier or inappropriate. Indeed, increasing a program's risk threshold may be the only viable technique for achieving "Safe Enough" if the vehicle's evolving design shows the original risk threshold to be unachievable. Ultimately, an appropriate risk threshold must strike a balance between what is *acceptable* and what is *achievable*.

34. Leveson N. Engineering a Safer World: Systems Thinking Applied to Safety. Cambridge, MA: MIT Press, 2011.
35. Feynman R. Personal observations on the reliability of the Shuttle. In: Report to the President By the Presidential Commission on the Space Shuttle Challenger Accident. Washington DC: Presidential Commission on the Space Shuttle Challenger Accident, 1986.
36. Broad W. High Risk of New Shuttle Disaster Leads NASA to Consider Options. The New York Times. 1989.
37. Aerospace Safety Advisory Panel. ASAP Public Meeting, Third Quarter 2010. 2010.
38. Hamlin T, Thigpen E, Kahn J, *et al.* Shuttle risk progression: Use of the shuttle probabilistic risk assessment (PRA) to show reliability growth. AIAA Space 2011 Conference & Exposition, Long Beach, CA, 2011.
39. NASA. ISS Crew Transportation and Services Requirements Document. NASA CCT-REQ-1130 Draft 4.0, 2011.

Address correspondence to:
Robert Ocampo
Aerospace Engineering Sciences
University of Colorado Boulder
429 UCB
Boulder, CO 80309

E-mail: robert.ocampo@colorado.edu