# CHALLENGES OF DETERMINING "SAFE ENOUGH" IN HUMAN SPACE FLIGHT

**Robert P. Ocampo[(1)] and David M. Klaus[(2)]**

*[(1)]University of Colorado Boulder, 429 UCB Boulder Colorado 80309 , robert.ocampo@colorado.edu*
*[(2)] University of Colorado Boulder, 429 UCB Boulder Colorado 80309 , klaus@colorado.edu*

## ABSTRACT

No spacecraft will ever be perfectly safe. Consequently, engineers must strive to design, develop, and operate spacecraft that are safe *enough*. But how safe is safe enough? The process of answering this question can be difficult and contentious, as the inherent uncertainties associated with defining and measuring "safe" are complicated by the subjective challenges of determining "enough." These complications, which include *uncertainty in terminology*, *subjectivity in the choice of metrics*, *uncertainty in the measurement itself*, and *subjectivity in the acceptance of the measurement*, must be eliminated, circumvented, or accounted for in order for "safe enough" to be objectively evaluated. This article describes these complications and presents recommendations in the hope that such discussion may help to facilitate the evaluation of "safe enough" in future spacecraft.

## 1. INTRODUCTION

Weeks before he would die in the tragic Apollo 1 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."*

To this day, space flight remains a "risky business". Recent accidents, including the loss of Orbital ATK'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?

These questions—and others like it—have surrounded the U.S. space program since its inception [1-10]. They remain relevant today because no comprehensive "right" answer exists: Different programs, flying different missions of various durations, may be willing to accept varying definitions and levels of risk, as well as varying degrees of uncertainty within the risk assessment [11]. The original Atlas booster was test flown 73 times before it was considered "safe enough" for crewed orbital flight; conversely, the Saturn V was test flown just *twice* before it sent humans to the moon.

Consequently, the question of "how safe is safe enough" must be continuously addressed, on both a program-by-program and flight-by-flight basis. This can be a difficult and contentious process, as the inherent uncertainties associated with defining and measuring "safe" are often complicated by the subjective challenges of establishing "enough." These complications—*uncertainty in terminology, subjectivity in the choice of metrics*, *uncertainty in the measurement itself*, and *subjectivity in the acceptance of the measurement*—are described below in the hope that such discussion may help to facilitate the evaluation of "safe enough" in future spacecraft.

## 2. COMPLICATIONS

### 2.1. Uncertainty in Terminology

The Aerospace Safety Advisory Panel (ASAP)—the independent group tasked with evaluating NASA's safety performance—has stated that the one of the primary obstacles to determining "safe enough" stems from the ambiguous use of the term "safe" in the English lexicon. In their 1978 annual report, 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*" [4].

This task of clarifying terminology is more complicated than a cursory dictionary definition would suggest, as the terms "safe" and "unsafe"—though *linguistically* antithetical—are often readily (if unknowingly) applied to the *same* spacecraft. Consider the case of two theoretical spacecraft: *Spacecraft A* and *Spacecraft B*. *Spacecraft A* has flown for 40 years without a single accident, while *Spacecraft B* suffered a catastrophic accident in its first flight. When asked to label these vehicles as "safe" or "unsafe", engineers readily characterize *Spacecraft A* as "safe" and *Spacecraft B* as "unsafe". However, both descriptions are derived from the same vehicle—in this case, the Soyuz spacecraft [12]. Such paradoxical labeling also appears to be present in descriptions of the Space Shuttle [11], suggesting the terms "safe" and "unsafe" are actually

context dependent, rather than (entirely) vehicle-specific.

These examples highlight just how difficult it can be to properly characterize crewed space systems. Labeling a spacecraft as "safe" is technically inaccurate, as no spacecraft can ever be free of catastrophic hazards (9, 11, 13-14]. Conversely, designating a spacecraft as "unsafe" implies that the vehicle cannot or should not be flown. These difficulties extend beyond a simple linguistic challenge, as using one of these terms over the other can potentially mean the difference between program viability and cancellation.

To resolve this dilemma, we have taken to describing "unsafe" in terms of *degrees*—articulated here and elsewhere as "risk" [11, 15]. Under this framework, if risk is determined to be sufficiently low, the spacecraft can be considered "safe enough." Conversely, if risk is determined to be unacceptably high, it can be rejected as "not safe enough"[1].

### 2.2. Subjectivity in the Choice of Metrics

Given these definitions, a metric for measuring spacecraft risk can now be selected. This metric serves to define the method of measurement, as well as the reference units for risk (e.g. number of successful launches, levels of failure tolerance, or performance in a flight readiness review [e.g. pass vs. fail]).

Selecting a metric to measure risk can be a difficult process. Unlike other physical variables, such as mass or length, risk cannot be measured empirically; it must be abstracted from the spacecraft and its interaction with the environment [16]. How this abstraction should proceed is ultimately a subjective choice (albeit one based on objective benefits). Should it based on the rate of successful launches? Or should it be quantified using probabilistic calculations? Or should some other metric be applied?

Apollo engineers were "deep[ly] and irreconciab[ly]" divided as to whether risk should be measured statistically or qualitatively [3]. Later Space Shuttle engineers faced a similar disagreement regarding the use of Probabilistic Risk Analysis (PRA) (*statistical*) vs. Failure Modes and Effects Analysis (FMEA) (*qualitative*) as the primary means of measuring risk post-*Challenger* [13, 17-19].

[1] Although there remains a certain degree of subjectivity to this choice of definitions (despite its heavy reliance on NASA terminology), they are employed here to ensure the discussion can proceed; without a set of firm definitions, the remaining complications (though largely independent of the definitions) cannot be effectively described.

These disputes were and remain prevalent because no metric can ever serve as a perfect proxy for risk. In order for statistical metrics to be reliable, the spacecraft must be tested thousands (if not millions) of times at both the component and system level. As some opponents have argued, this time may be better spent "searching for design flaws than mindlessly running tests" [3]. On the other hand, measuring risk based on qualitative engineering judgments tends to produce widely divergent assessments of risk. Prior to the *Challenger* accident, experts qualitatively estimated the probability of a catastrophic Space Shuttle accident to range between 1 in 100 and 1 in 100,000—*a range of three orders of magnitude* [17].

Even relatively simple metrics, such as dividing the number of successful launches by the number of total launches, cannot serve as perfect proxies for risk. 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" [20].

To date, a consensus method for evaluating risk remains undefined in the United States. NASA currently places special emphasis on prospective metrics, such as Probabilistic Risk Assessment (PRA), to measure risk in its Commercial Crew Program (CCP) [21]. This contrasts with FAA regulations, which specify the use of retrospective, actuarial metrics to define vehicle safety records [22].

### 2.3. Uncertainty in the Measurement Itself

The choice of a metric is further complicated by the fact that few (if any metrics) can quantify risk with perfect precision. Even something as simple as counting redundant parts within a system—an approach loosely (and more qualitatively) employed by the Space Shuttle Program in its early years under the auspices of the Critical Items List (CIL) [13]—can generate uncertainty. The Space Shuttle Solid Rocket Booster (SRB) O-rings 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 to catastrophic failure) in 1982, when engineers realized that leakage of the primary O-ring during certain phases of launch was actually a single-point failure [23]. Notably, this classification change occurred in the absence of any modifications to the SRB design, suggesting that even simple metrics for measuring risk can have uncertainty associated with their measurements.

This uncertainty must be accounted for to ensure evaluations of "safe enough" are statistically

appropriate. Consequently, certain metrics explicitly list an uncertainty component in their measured values. Probabilistic Risk Assessment (PRA), for example, generates both a mean estimate of risk *and* an estimate of risk uncertainty.

### 2.4. Subjectivity in the Acceptance of the Measurement

Once a method for measuring risk has been selected, a threshold value (either quantitative or dichotomous) can be assigned to "safe enough" (e.g. ***18*** successful launches, ***3*** levels of failure tolerance, Probability of Loss of Crew less than ***1/200***, ***successful completion*** of a flight readiness review). If a spacecraft meets (or in certain cases, exceeds) the threshold value, it can be considered "safe enough"; if it does not meet this threshold, it can be rejected as "not safe enough" [11].

This threshold value must balance what is *achievable* given the program's budget, schedule, and engineering capabilities with what is desirable (or *acceptable*) from a programmatic or personal standpoint [7, 11]. Identifying such a value is no simple or arbitrary task. Predicting *achievable* risk requires the use of expert opinion during the initial stages of program development [9]. As described in the Space Shuttle example above, such opinions can vary widely, and even then may not encompass the "true" risk of the system.

Determining what is *acceptable* in terms of risk is also "far from straightforward", as anticipated mission benefits must be shown to demonstrably outweigh the potential for mishap [24]. Given that the "weight" of each benefit naturally varies from individual to individual (and program to program), this process is largely subjective. As such, it is amenable to rejection *ex post facto*. Consider a theoretical spacecraft that meets its assigned risk threshold with perfect certainty. Even though such a spacecraft is "safe enough" by the letter of the law, it would be hard to argue as such in the event of a catastrophic accident.

## 3. RECOMMENDATIONS

Developing techniques to eliminate, account for, or circumvent these complications is critical to the consistent and unequivocal evaluation of "safe enough". Although a detailed description of these techniques is beyond the scope of this article, several general recommendations are described below:

- Define all relevant terminology (e.g. "safe", "unsafe", "safe enough", and "risk") during program startup, and strictly adhere to these definitions throughout the program's development and operations. This may serve to minimize confusion as to what is "safe" versus what is "safe *enough*."

- Select a risk metric that either minimizes or accounts for measurement uncertainty. This helps to ensure that "safe enough" can be reliably evaluated (to a pre-determined level of statistical certainty) despite any potential unknowns in the vehicle's performance or design.

- Establish an acceptable risk threshold based on what is realistically achievable given technical, budget, and schedule constraints. Expert opinion and risk heuristics should be used early in the design process to coarsely predict achievable risk. If risk is determined to be acceptably low, the program can continue forward; conversely, if risk is determined to be unacceptably high, then the program can be restructured (or cancelled) before significant resources are committed to a specific design or operational paradigm [25].

## 4. CONCLUSION

Answering the question "how safe is safe enough?" is a difficult and contentious process, impeded by elements of subjectivity and uncertainty. Despite these complications, the question of "safe enough" should be placed at the forefront of all human space flight programs, as the answer will determine the level of risk that is considered acceptable by the stakeholders and also serve as a benchmark for assessing whether or not it is achieved.

To Gus Grissom and his fellow astronauts, the "conquest of space [was] worth the risk of life." We must continuously ask: "is it still?"

## 5. ACKNOWLEDGMENTS

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.

## 6. REFERENCES

1. 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*.

2. Hacker, B. C., & Grimwood, J. M. (1977). On the Shoulders of Titans: A History of Project Gemini:'Spirit of'76. *NASA, Rept. SP-4203*, 265-298.

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

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

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

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

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

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

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

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

11. Ocampo, R. P., & Klaus, D. M. (2016). A Quantitative Framework for Defining "How Safe is Safe Enough" in Crewed Spacecraft. *New Space*. 4(2): 75-82

12. Ocampo, R.P. (2016). *Unpublished data*.

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

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

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

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

17. 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.

18. 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).

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

20. Ocampo, R. P., & Klaus, D. M. (*2016*). Comparing the Relative Risk of Space Flight to Terrestrial Modes of Transportation and Adventure Sport Activities. *New Space*. 4(3): 190-197.

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

22. Operator informing space flight participant of risk, 14 C.F.R. § 460.45 (2016).

23. Presidential Commission On Space Shuttle Challenger, & Rogers, W. P. (1986). Report of the Presidential Commission on the Space Shuttle Challenger Accident.

24. Aerospace Safety Advisory Panel. (2013). Aerospace Safety Advisory Panel Annual Report for 2012. *Washington, DC*.

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