An Expected Value Approach to "As Safe as Reasonably Practicable"

Everett H.C.[a]
[a]Idaho National Laboratory, Idaho Falls, USA

Abstract: Being as safe as reasonably practicable (ASARP) is a fundamental principle of adequate safety at NASA. A system is ASARP if an incremental improvement in safety would require a disproportionate deterioration of system performance in other areas. This special treatment of safety relative to other objectives is based on the idea that life is precious, and therefore safety should not be traded away in pursuit of other objectives unless absolutely necessary. This view of safety is at odds with standard decision analysis, which assigns a definite value to each possible outcome in the decision space, including safety outcomes, and identifies the decision alternative that maximizes expected value. Being ASARP doesn't enter into the analysis. This paper presents an approach to ASARP that charts a middle path, working within an expected-value framework while still respecting the incommensurability of safety. It is based on inverting the expected value equation to ask, "Which decision alternative supports the highest valuation of the lives put at risk, while still being worth implementing?" The alternative so identified is deemed ASARP.

Keywords: NASA, Safety, Expected Value, Expected Utility, ASARP

1. INTRODUCTION

Being as safe as reasonably practicable (ASARP) is a fundamental principle of adequate safety at NASA. As discussed in the NASA System Safety Handbook, the ASARP region of a trade space contains those alternatives whose safety performance is as high as can be achieved without resulting in intolerable performance in one or more other domains. In practice, however, identifying an ASARP alternative from some set of alternatives is challenging, since on the one hand the concept of ASARP is based on safety being incommensurable with other objectives, while on the other hand any solution that puts people at risk implies some valuation, stated or not, of human life. In many industries, an explicit valuation of human life is made, and standard decision analysis is conducted with a goal of maximizing expected value. Unfortunately, this approach is antithetical to the ASARP principle. This paper presents a simple expected-value approach to identifying an ASARP alternative that does not depend on an a priori valuation of human life and is conceptually respectful of the incommensurability of safety. It is based on inverting the expected value equation to ask, "What is the most value that can be put at risk on a mission and still be worth it?"

2. ASARP BACKGROUND

The term "ASARP" was introduced into the NASA lexicon in 2011 in the NASA System Safety Handbook [1] as one of two fundamental principles of adequate safety. As shown Figure 1 from [2], an adequately safe system is one that not only meets some minimum tolerable level of safety below which the system would be considered unsafe, but is also as safe as reasonably practicable regardless of its absolute level of safety, on the grounds that NASA has an ethical obligation to minimize the potential for harm in the conduct of its missions. The origin of this obligation lies in the closely held belief that life is in some sense precious beyond measure and should never be put at unnecessary risk. In other words, safety is incommensurable with other objectives, such as those related to profit, time, political dominance, or technology development, and to trade it against such pursuits is to some degree an affront to human dignity. Unfortunately, safety risk is an inherent aspect of space exploration, and the only way to avoid it completely would be to abandon the enterprise. ASARP represents an effort to find a middle ground where safety is given top priority within the context of accomplishing missions that necessarily put safety at risk.

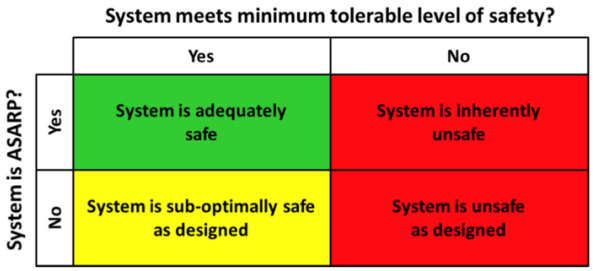

Figure 1. Adequate Safety (Reproduced from [2])

As discussed in [1], "A determination that a system is ASARP entails weighing its safety performance against the sacrifice needed to further improve it. The system is ASARP if an incremental improvement in safety would require a disproportionate deterioration of system performance in other areas." This is illustrated in Figure 2, which shows a space of identified alternative system configurations in terms of their safety performance versus their performance in other areas. The configurations that are not categorically inferior to other alternatives are those that lie on the efficient frontier. The ASARP region of the space of alternatives is therefore the region of the efficient frontier where safety is maximized to the point where further increases in safety would begin to result in intolerable performance in other areas such as cost, schedule, or technical performance.

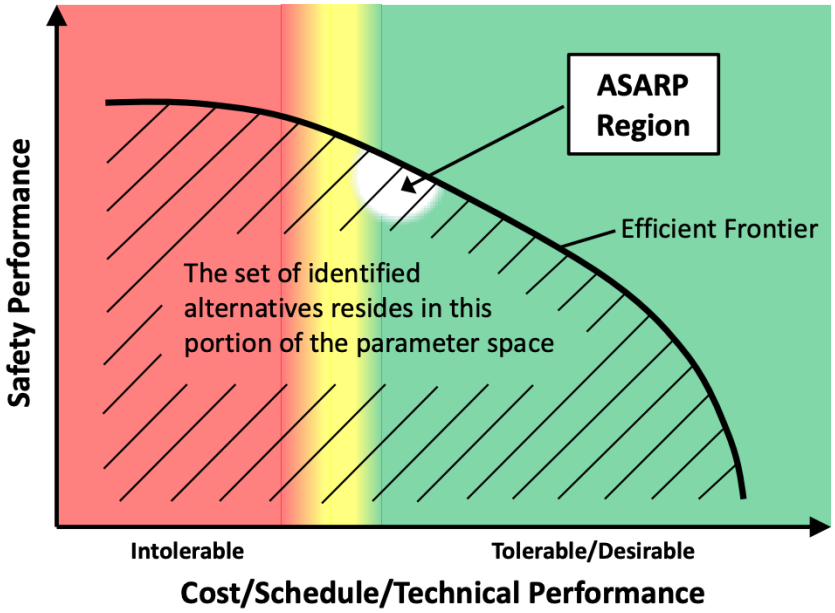

Figure 2. The ASARP Region of a Space of Alternatives (Reproduced from [1])

3. CONFLICT BETWEEN ASARP AND DECISION-ANALYTICAL METHODS

Modern decision analysis is largely based on the assignment, to each possible outcome of a decision, of a value or a utility, and the goal of the analysis is to determine which decision alternative maximizes the expected value or the expected utility. The use of expectation is needed to account for uncertainty in the specific outcome of an alternative. For example, in choosing between a high cost, high reliability system and low cost, low reliability system, the expected values of each alternative are calculated from their

respective probabilities of success, their respective costs, and the value of the mission return being pursued, as shown in the following equations:

Expected value of high-reliability system:

$$E[V]_{HR} = P_{S\_HR}R - C_{HR} \quad (1)$$

Expected value of low-reliability system:

$$E[V]_{LR} = P_{S\_LR}R - C_{LR} \quad (2)$$

Where:

E[V] = Expected value
P_S = Probability of success
R = Value of the mission return
C = System cost

If $E[V]_{HR} > E[V]_{LR}$, then the high-reliability system is preferred. If $E[V]_{HR} < E[V]_{LR}$, then the low-reliability system is preferred. Importantly, the calculation of expected value requires that all items of value can be expressed in terms of a common unit of measure, i.e., that they can all be "dollarized."

If each of the above alternatives involves a life safety risk where fatalities and mission failure are linked, then the above equations become:

Expected value of high-reliability system:

$$E[V]_{HR} = P_{S\_HR}R - C_{HR} - (1 - P_{S\_HR})L \quad (3)$$

Expected value of low-reliability system:

$$E[V]_{LR} = P_{S\_LR}R - C_{LR} - (1 - P_{S\_LR})L \quad (4)$$

Where:

L = Value of the lives put at risk

Equations 3 and 4 require a valuation of the lives put at risk, and indeed this is often done in commercial risk management (e.g., in the automobile industry) and regulatory analysis. For example, the U.S. Department of Health and Human Services set the 2026 value per statistical life (VSL) at between $6.2 million and $20.3 million, with a central estimate of $13.4 million [3].

Equation 3 or 4 can be inverted with respect to the probability of success, as shown in Equation 5, and plotted against cost, as shown in Figure 3.

Probability of success:

$$P_S = [1/(R + L)]C + [(E[V] + L)/(R + L)] \quad (5)$$

Figure 3 shows lines of constant expected value overlayed on a space of alternatives where the ASARP region is that part of the efficient frontier that pushes up against a cost constraint, consistent with [1]. The

slope of each line is the same, namely [1/(R + L)], and the line for E[V] = 0 intersects with the point (C = R, P_S = 1) in the upper right corner of the figure. The system that maximizes expected value is at the point on the efficient frontier that intersects with a tangent line of constant expected value. Movement away from this point along the efficient frontier (or into the interior of the space of alternatives) necessarily decreases the expected value. Movement up and to the right decreases it because further increases in safety aren't worth the cost. Movement down and to the left decreases it because the cost savings aren't worth the additional risk.

Decision-analytical methods such as the maximization of expected value do not involve the concept of ASARP. Instead, given the "dollarization" of all relevant outcomes and the determination of all relevant probabilities, the best course of action is simply the one with the maximum value of E[V]. Such methods reject the notion that safety is incommensurable, and instead place a potentially high, but definite, value on the lives put at risk as the basis for trading safety against other objectives.

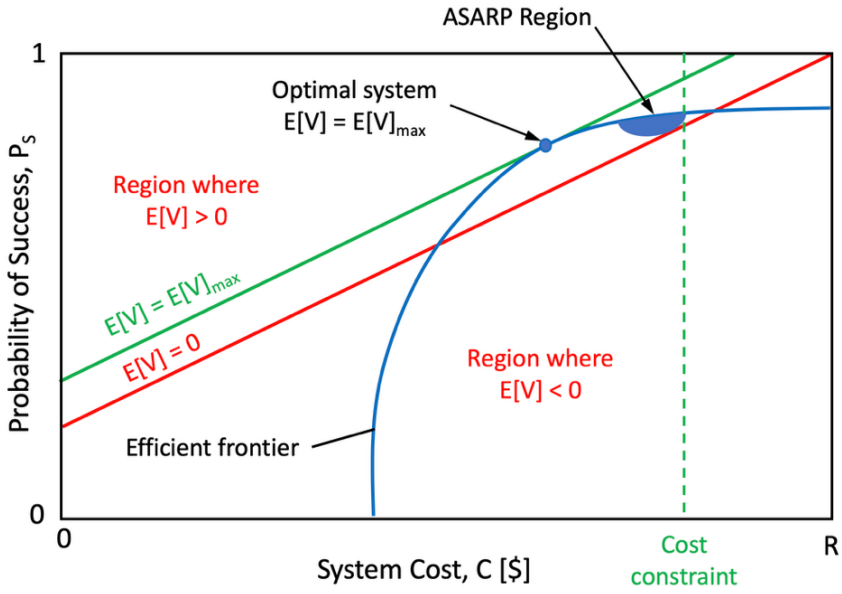

Figure 3. Optimal System for a Given Value of Lives Put at Risk

While the ASARP region as described in [1] is largely determined by pushing against tolerable limits and is stable insofar as the tolerable limits are stable, the system that maximizes expected utility depends on the value of the lives put at risk, and different valuations will yield different solutions. Specifically, higher valuations will decrease the slope of the lines in Figure 3, which will shift the tangent point up and to the right, whereas lower valuations will increase its slope and shift the tangent point down and to the left. This renders the method sensitive to changes in valuation and vulnerable to disagreements about the appropriate valuation.

4. ASARP IN AN EXPECTED VALUE CONTEXT

The key insight of this paper is that for each alternative there is a valuation of the lives put at risk above which the risk of fatalities outweighs the potential benefit of the mission. Given this, it is possible to find the system (or possibly systems) associated with the highest such valuation. This system can be considered ASARP because it's the only one that can rationally be entrusted with lives of that value from an expected value standpoint. The selection of any other system on the efficient frontier (or into the interior of the space of alternatives) will result in a mission that is not worth conducting given the cost, risk, and potential return.

For the above example, Equations 3 and 4 can be inverted with respect to the lives put at risk to yield:

Value of the lives put at risk:

$$L = (P_S R - C - E[V])/(1 - P_S) \quad (6)$$

For a given system, the maximum allowable value for the lives put at risk is at the tipping point between the mission being worth conducting and not being worth conducting. In other words, it is at the value where E[V] = 0, such that:

Maximum value of the lives put at risk for alternative *i*:

$$L_i = (P_{S\_i} R - C_i)/(1 - P_{S\_i}) \quad (7)$$

The system with the highest value of L given E[V] = 0 (i.e., L_{Max}) is at the point on the efficient frontier that contacts the tangent line containing the point (C = R, P_S = 1). The situation is shown graphically in Figure 4, where the system for which L = L_{Max} is identified as the "Expected-Value-Based ASARP System." This is the only system that can be rationally entrusted with lives of value L_{Max}.

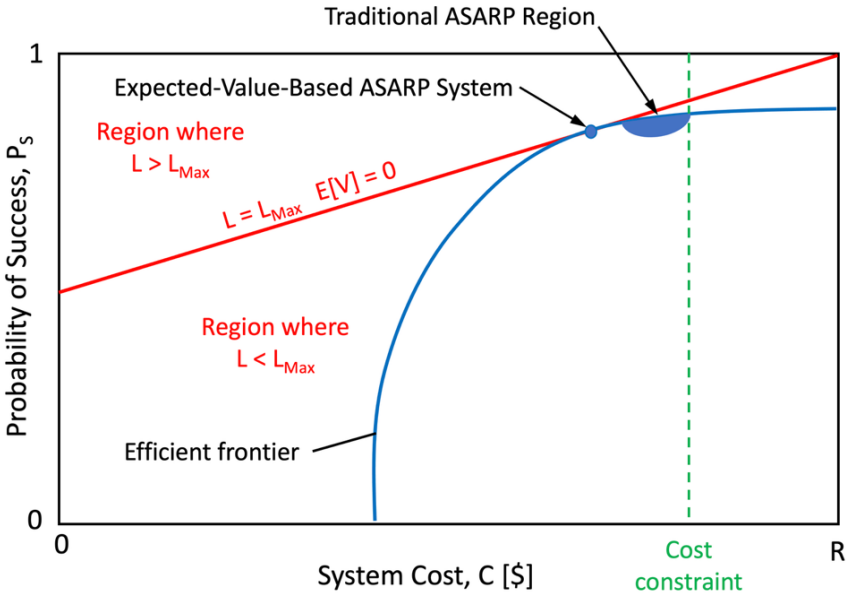

Figure 4. ASARP in an Expected Value Context

Although the method calculates a valuation L_{Max} for the lives put at risk, it does not assert that their value is indeed L_{Max}. Rather, L_{Max} is merely an artifact of the method. Moreover, the method itself is silent on the question of whether or not the safety risk of the mission is acceptable. This is consistent with the concept of adequate safety as presented in [1] and illustrated in Figure 1, where the question of the tolerability of a given level of safety risk is separate from the question of whether or not the system is ASARP. Instead, the method reflects the idea that if life is precious beyond measure, but at the same time the mission return is worth the risk required to obtain it, then the system used should be the one that can be entrusted with the highest possible hypothetical valuation of the lives put at risk that still supports a rational decision to make the attempt.

The expected-value-based ASARP system of Figure 4 is identical to the optimal system of Figure 3 in the case where L is explicitly valued at L_{Max}. For all other valuations of $L < L_{Max}$, the slope of the lines in Figure 3 are greater than for $L = L_{Max}$, and therefore the probability of success is lower. (For $L > L_{Max}$, there is no system where $E[V] > 0$.) Therefore, not only is the expected-value-based ASARP system the one that can be entrusted with the highest valuation of the lives put at risk, but from among those systems whose use is conceivably justifiable on expected value grounds, it is the one with the lowest probability of loss.

5. THE SPACE OF ALTERNATIVES

As discussed above, the ASARP principle is separate and independent from any safety risk tolerances that may be levied on the mission to define thresholds of acceptable safety risk. Instead, it is concerned with the safety of each alternative relative to that of the other alternatives that could have been selected. As such, a claim of ASARP is predicated on a reasonable, sustained, and proactive search for the safest practical alternatives throughout the entirety of the activity life cycle (e.g., from Pre-Phase A: Concept Studies, through Phase F: Closeout). Each decision that significantly affects safety should consider a sufficiently broad set of alternatives. In particular, a minimum condition for ASARP is the consideration of applicable established good system safety and safety management practices in decision-making, such as the use of the best available technology (BAT). Established practice is typically captured in consensus technical and process standards and can be focused on very specific details of design, manufacture, analysis, management, operation, etc. [4].

The consideration of best available technology and practice, rather than the requiring of it, is consistent with recent recommendations in other contexts such as the nuclear power industry. Both the Idaho National Laboratory and the UK Nuclear Regulatory Review argue, with respect to the closely-related As Low As Reasonably Achievable (ALARA) and As Low As Reasonably Practicable (ALARP) principles, that the requirement to use ‘relevant good practice’, “has become far too prescriptive in the pursuit of risk reduction,” and that “the effect of stepping up measures until they are far beyond the ‘reasonable’ definition” leads to “excessive nugatory effort and the stifling of innovation” [5].

The expected-value-based ASARP method presented in this paper entails the consideration of relevant good practice such as BAT, but also allows for the consideration and use of novel practices. This is particularly important in the current spaceflight environment, which not only routinely involves technology development, but which is increasingly led by pioneering commercial providers.

6. GENERALIZATION OF THE METHOD

The expected value example presented above applies to the simple situation where life safety and mission success are linked together, where dollar value can stand in for utility, and where the value of the lives put at risk can be easily hypothesized. In more complex situations, it might be the case that the risk to mission success is somewhat decoupled from safety risk, a utility-based rather than value-based formulation is appropriate (e.g., due to non-linearities), or the valuation of the entities put at risk (whether lives, other precious assets, or both) is complicated by the differences among them.

The ASARP method presented here can be readily adapted to such situations with the following caveats:

- The example above implicitly assumed that the expected value of the “no mission” alternative is zero, so that the tipping point between conducting the mission and not conducting the mission can be found by setting $E[V] = 0$. In the general case, the expected utility, $E[U]_0$, of the “no mission”

alternative defiines the tipping point but is not necessarily zero[1], so that the equation for finding the maximum value of L for alternative *i* is:

Tipping point for the general case: $E[U]_i = E[U]_0$ (8)

In a particular application, Equation 8 might not be invertible with respect to L, as Equation 6 is, and numerical methods might be needed.

- The example above assigns a value R to the mission return. However, in a real-world application it is not necessarily easy to assign a definite value to R, particularly because much of what NASA does is discovery, and it can be challenging to know what the returns will be or even what the space of possible returns looks like. One approach to managing this uncertainty is to characterize the ASARP alternatives parametrically as a function of R. For a single-objective mission the parameter space would be one-dimensional. For a multiple-objective mission the parameter space would be multi-dimensional. In any case, understanding whether or how the ASARP alternative changes as a function of R would likely provide valuable insight into choosing which alternative to select.

- Although the method only uses L in a hypothetical context, if an alternative puts multiple entities of different types at risk, it might be necessary to assign relative valuations to each type. This would be the case when the selection of the ASARP alternative requires trading safety risk to one type of entity against safety risk to another type of entity. The assignment of relative valuations enables $E[U]_i$ to be a function of a single quantity, L, representing the (hypothetical) total value of the entities put at risk. In cases where relative valuations are difficult to make, a parametric approach similar to the one described above for R could be used.

7. CONCLUSION

This paper presents a simple and straightforward approach to the ASARP principle that is compatible with expected-value-based decision analysis while respecting the incommensurability of safety. It is true to the spirit of ASARP in that it finds the alternative that among all the alternatives considered is compatible with the highest valuation of the entities put at risk.

The method is notable for what it doesn't entail:

- It doesn't entail pursuing safety to the edge of tolerable performance (and risk) in other areas.

- It doesn't entail the rote use of BAT or established best practice absent an analytically based justification for it.

Because the method depends only on the space of alternatives and the valuation of objectives, it can be used early in the life cycle (e.g., before cost constraints have been established). Indeed, the identified ASARP alternative can be used to scope the effort, including cost estimates and development timeframes. Moreover, it provides an objective technical basis for pushback against any "ASARP creep" that might otherwise occur over the life cycle.

[1] For example, it is common to assign a utility of one to the best possible outcome and a utility of zero to the worst. Assuming that the worst outcome is associated with a failed mission, the "no mission" alternative would presumably have a utility greater than zero (but hopefully less than one).

Acknowledgements

The author would like to thank the NASA Office of Safety and Mission Assurance (OSMA) for their sponsorship of this paper through Strategic Partnership Project (SPP) #01717 - Development and Implementation of Risk Management and System Safety Framework at NASA. The views and opinions of the author expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

References

[1] NASA/SP-2010-580, NASA System Safety Handbook Volume 1: System Safety Framework and Concepts for Implementation. NASA. 2011.
[2] NASA/SP-2014-612, NASA System Safety Handbook Volume 2: System Safety Concepts, Guidelines, and Implementation Examples. NASA. 2014.
[3] A. Kearsley, "HHS Standard Values for Regulatory Analysis, 2024." HHS. 2024.
[4] NASA/SP-20250002210, Guidance on the Implementation of an Objectives-Driven, Risk-Informed, and Case-Assured Framework for Safety and Mission Success, NASA. 2025.
[5] "What Future for the ALARA Principle?", NEI. February 17, 2026.