Factors that Influence Acceptance of New Aerospace Risk Assessment Techniques

M.L. Cummings[1]
Duke University, Durham, NC, 27708, USA

With the rise of new space missions such as NASA's Dragonfly that will include a nuclear-powered rotorcraft for exploring Saturn's icy moon, Titan, as well as deep space missions that cannot rely on solar power, new risk assessment strategies are needed that balance the need for safety against the impracticality of lifetime testing. Risk assessments for existing nuclear technologies and, more broadly, safety-critical systems can widely vary in approaches and outcomes, with significant cultural influences. Some safety-critical systems like driverless cars and surgical robots have been authorized to operate with little or highly questionable risk assessments. Others using well-established probabilistic-based risk assessment methods such as those used for Stirling-based convertors for Radioisotope Thermoelectric Generators have struggled to convince relevant agencies that risk is acceptable. Case studies show that risk-averse oversight groups tend to rely more on concepts of heritage for technology risk assessments, which is the existence of either observed data from actual or similar operational systems. While reliance in heritage may reduce risk, it can result in incremental, evolutionary technologies instead of revolutionary ones, in effect stifling innovation. This notion of heritage can also lead to a misperception of acceptable risk under the guise of a regulatory concept termed equivalence, which allows new technologies to be fielded based on erroneous technical comparisons, much like the Boeing 737 MAX. More work is needed to understand the interplay between probabilistic-based risk assessments, notions of equivalence and heritage, and the culture of oversight agencies.

I. Nomenclature

A, B	=	material constants
CTF	=	cycles to failure
N_f	=	number of cycles-to-failure
σ_α	=	fully reversed fatigue strength

II. Introduction

It has long been the ambition of humanity to stretch the outer limits of our understanding of the solar system and galaxies beyond. To meet these ambitious goals, NASA and other space agencies will need to develop more innovative technologies capable of overcoming challenges inherent to these missions, including human adaptability to long duration missions, the need for more autonomous spacecraft and terrestrial exploration devices, as well as long-duration power sources that do not rely on solar power. Risk is an inherent part of space missions, and space agencies must determine not whether to take risks, but rather where and how to take them so they can be managed more effectively [1].

To this end, NASA has been developing next-generation space power systems such as the dynamic Radioisotope power system (DRPS). While NASA has relied on static power systems for nearly 70 years in the form of radioisotope thermoelectric generators (RTG), a DRPS attempts to improve RTG fuel efficiency and decrease mass through Stirling and Brayton cycle heat engine technology as the basis of the converters. However, one risk-related issue that arises in the development of such systems is how to develop a testing program that can provide acceptable risk estimates for technologies that cannot be fully tested either in their operational domains or for the expected life of the system.

[1] Professor, Electrical and Computer Engineering, AIAA Fellow.

To address the need for improved risk assessment approaches and technologies, NASA encourages agencies to follow established probabilistic risk assessment (PRA) approaches or even create new risk analysis techniques [2-4]. PRA techniques are well-established across a number of NASA applications such as assessing the risk and consequences of a spacecraft's propellant leak, launch failure, and human-system interactions [5], but also in other applications like assessing health-patient safety [6] and in the assessment of nuclear reactor design weaknesses, where PRA was first devised [7].

However, there are issues with PRA, particularly with new, innovative technologies like DRPS. The performance lifetimes of these systems are expected to be long at 17 years [8]. Current testing and PRA approaches are very expensive and extremely time intensive. For example, traditional statistical methods used to validate system reliability of .9 at 15 years with a confidence of 90% for an RPS would take 2.9M hours of system demonstration without failure. Given these constraints, the Radioisotope Power Systems Project (RPSP) Office and the Johns Hopkins University Applied Physics Laboratory (APL) developed a new derivative PRA process called RILT (Risk Informed Lifetime Modeling) process to determine the target, nature, and extent of testing required for the demonstration of acceptable RPS reliability [8].

Unfortunately, while the computation of risk may seem like an objective process, PRAs often carry significant subjective elements, not the least of which is interpreting probabilistic results and determining safe thresholds, which are notoriously difficult decisions for humans to make [9-11]. Given the subjective aspect of risk assessment, it is not uncommon for various groups to debate whether a given risk assessment approach is correct or the best candidate for a particular application. For instance, alternatives to PRA include the examination of trade studies for relevant technologies and convening technical peer reviews with those not immediately involved with the project under evaluation [12].

Similarly, RILT is a relatively new risk assessment technique and may not be readily accepted across NASA or its partners and subcontractors. DRPS has yet to fly on any space mission, and without a convincing risk analysis, it will not be flown. To further examine these issues, this paper investigates those factors that influence the acceptance of a new risk analysis technique like RILT, what issues can arise while introducing a new risk analysis technique, and what can be done to mitigate these issues to allow for the cautious development of innovative technologies.

III. Risk Informed Lifetime Modeling

RILT is a hybrid risk assessment methodology that combines physics of failure modeling with traditional probabilistic risk assessment modeling, informed by past flight data, if available, and expert judgement [8]. Physics of failure modeling is an approach that assesses a system's reliability by simulating models of failure based on well-documented phenomena such as fatigue, fracture, wear, and corrosion. According to RILT's developers, "by providing a failure model through which millions of hours of testing can be simulated, RILT, specifically its inherent Bayesian treatment of evidence, allows for evaluation of effects of pseudo-life test data on life predictions. This means that parameters that drive uncertainty in the simulation environment can readily be identified and targeted for actual life-testing [8]."

To illustrate an application of RILT, take the need to determine the lifetime performance of piston flexure bearings within a Stirling generator [8]. In the case of DRPS, Stirling generators use the heat of decaying radioisotopes and low temperatures of space to expand and compress gases to drive an alternator. The flexure bearing supports the longitudinal movement of the Stirling converter linear alternator while limiting the alternator's radial motion. One likely failure mode evaluated through RILT is the rate of fatigue until failure of the flexure bearings resulting from use over a system's lifetime.

The equation used to model the lifetime performance of the flexure bearing under various levels of stress, that is the number of cycles performed by the bearing until failure, is presented in Eqn. 1, known as the S-N curve, where the stress amplitude, σ_α, or the reversed fatigue strength of the bearing is directly related to the number of cycles-to-failure (N_f) as well as material constants A and B.

$$\sigma_\alpha = A N_f^B \tag{1}$$

Equation 1 can be linearized and Bayesian regression applied to estimate the distributions of A and B material constants. Reorganized to make the cycles-to-failure the dependent variable and the fatigue strength of the material the independent variable, Eqn. 1 becomes:

$$\log N_f = \frac{1}{B} \log \sigma_\alpha - \frac{1}{B} \log A \tag{2}$$

Cycles-to-failure for metal used in the bearing can then be computed as a function of stress amplitude (Table 1), and probability density functions can be computed for A and B, which are parameters representing material constants.

In this way, RILT estimates the maximum stress that the flexure bearings can withstand during the lifetime of the generator, as well as the uncertainty associated with these parameters.

Table 1: Example S-N Data

Stress Amplitude, MPa	*Cycles-to-failure*
948	222
834	992
703	6004
631	14130
579	43860
524	132150

Table 2: Summary statistics of model coefficients and cycles-to-failure (CTF) at 500 MPa

Parameter	*Mean*	*SD*
A	1.588e+03	1.803e+01
B	-9.453e-02	1.2473-03
CTF	2.037e+05	1.081e+04

The purpose of RILT is to use Bayesian probabilistic risk assessments and accelerated testing data in order to identify and target those variables and operational regimes that are primary risk drivers. Accelerated testing is a process by which designers attempt to simulate, over a relatively short period of time, the conditions that a part, component, assembly, or an entire system will experience throughout its anticipated service life. It is an element of reliability engineering meant to discover and eliminate failure modes, and is commonly used across many industries, particularly for electronic components [13] and materials [14].

Accelerated testing provides data for risk assessment, and NASA has long used such testing for space vehicle components [15, 16]. Accelerated testing informs, but does not replace probabilistic risk assessment, e.g., [17]. Data from accelerated testing allows managers to build appropriate levels of trust and confidence in such models, which then feeds a risk assessment which considers data and input from multiple sources.

Figure 1 illustrates the relationship between RILT and accelerated testing in the context of data fidelity and system abstraction. In an ideal world, the best risk assessment for any technology would be based on real data taken from actual parts and components, working in an operational integrated system. Given the complexities surrounding nuclear space power including cost and remote operations, it is simply not possible to perform such an ideal risk assessment. As depicted in Fig. 1, if data from an identical system is not available, often risk assessments are based on data from similar systems [18].

The existence of either observed data from actual or similar operational systems is a concept NASA refers to as "heritage," depicted in Fig. 1. Proposals for new systems that can point to heritage in parts, components, and systems are typically seen to be less risky and more likely to succeed [19]. In domains outside of NASA, a related notion of equivalence between two similar technologies often forms the basis for regulatory certifications [20].

For systems, components, or parts without such heritage in an operational setting, more data is needed to accurately assess risk. Accelerated testing, as discussed previously, can help fill this gap and as depicted in Fig. 1, can span testing across parts, components, and even potentially systems. Data gathered from physical testing forms physics-based models that can be used to estimate future states through probabilistic models. The fidelity of such estimates depends on the data gathered and how well that data can generalize to unseen or predicted states. Such approaches have a significant limitation in that there is typically an assumption of linearity. It is possible that underlying behavior in such systems is non-linear, and thus any model with linearity assumptions, like RILT, will be inherently flawed.

Based on publicly-available literature, in Fig. 1 RILT is depicted as having significant overlap with accelerated testing, which is a stated characteristic of this approach [8]. RILT uses similar physics-based modeling approaches as accelerated testing, a commonly-accepted risk assessment strategy. However, a hallmark of RILT is that it relies heavily on probabilistic data for performance estimates at the part and component level, i.e., its *"inherent Bayesian treatment of evidence* [8]," which then allows for more targeted testing for the larger system.

In Fig. 1, RILT is depicted as straddling the lower data fidelity levels while spanning system abstractions from concepts through components. RILT can rely on some actual observed data, but as is depicted in public literature, it is a tool to identify critical testing regimes using significant probabilistic analyses. However, given the propensity of decision makers to trust observed data from actual or similar operational systems, i.e., heritage, it is very likely that the lack of such data lowers the trust in RILT results due to the higher uncertainty in predictions.

One strength of RILT that can easily be overlooked is that of all the risk assessment data generation methods depicted in Fig. 1, it is the only one designed to capture data for examining conceptual system designs. Because RILT can capture a wide array of probabilistic scenarios, it has tremendous value in allowing for the exploration of potential

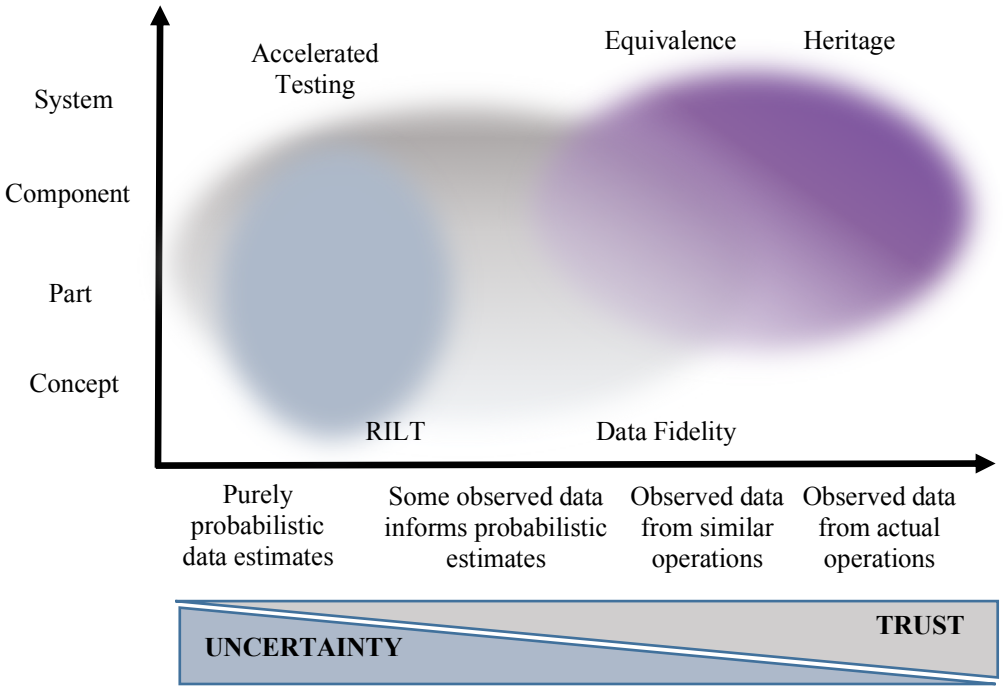


Fig. 1: The risk assessment relationship between system abstraction, data fidelity, and testing.

risk for various concepts of operations informed by real data where such data exists. Thus, it allows for the mixing of observed and estimated data to allow for state space exploration not available through the use of other methods.

While RILT and other more probabilistic approaches to risk assessment may be seen as less data-driven and problematic, it is simply often not possible to gather the gold standard of data from actual or similar operations. Indeed, this is a significant problem for NASA managers who would like to field futuristic space explorations missions where current power systems simply are not adequate [21], including the new Dragonfly mission. This conundrum highlights a classic problem in engineering in that major breakthroughs in engineering cannot happen without taking some risk [22], and often agencies tend to avoid risk instead of managing it, which ultimately suppresses innovation [23].

To this end, the next section examines both current and futuristic systems in terms of risk assessment to further examine how NASA and other agencies think about risk and how to mitigate negative outcomes in the adoption of new risk assessment strategies.

IV. Risk Assessment Comparison for Actual and Prototype Technologies

To better understand how different approaches to risk assessment can affect the final decision to deploy a technology, Fig. 2 illustrates assessment strategies for various operational and prototype technologies, and the associated system abstraction level. Each technology will be discussed in more depth in the next sections, but they include a mix of NASA and other technologies to provide perspective.

A. Deployed Systems

As discussed previously, the best risk assessment data for proposed technologies comes from systems that currently exist, or are very similar. The case of the Mars InSight lander in the upper right of Fig. 2 demonstrates how this notion of heritage can influence NASA decisions. InSight (Interior Exploration using Seismic Investigations, Geodesy and Heat Transport) is a JPL lander designed to study the crust, mantle, and core of Mars. It was one of three designs proposed under the 2010 Discovery competition. The other two were the JHU-APL Titan Mare Explorer (TiME) for exploration of a large methane-ethane sea on Titan and Goddard Space Flight Center's Comet Hopper (CHopper), which would have orbited and repeatedly landed on Comet Wirtanen.

These three systems were partially developed through an initial round of funding to determine which would go on to full development. Both TiME and Chopper proposals included a DRPS system, the Advanced Stirling Radioisotope Generator (ASRG), which was incentivized by NASA in the competition. NASA will often incentivize particular technologies in order to advance their Technology Readiness Level (TRL). Despite this incentivization, the solar-

powered InSight was the winner of this competition, in part because the risk was lower as solar-powered flight has significant heritage. At that same time, the ASRG faced significant technical challenges [24].

While solar power has high heritage as a power system for space vehicles, nuclear-powered RTG systems have flown in space many times. In 2004, JPL successfully proposed a derivative of an older RTG design in the form of the Multi-Mission Radioisotope Thermoelectric Generator (MMRTG) to power the Mars Science Laboratory Curiosity rover. This MMRTG was selected over a Stirling-based DRPS, which in theory would have been more efficient. The MMRTG was selected primarily due to its heritage with similar thermoelectric couple designs and materials from earlier RTGs that flew on Viking and Pioneer missions. Recall that a Stirling-based system introduces a new moving part discussed in the earlier RILT example, so it is considered a more radical departure from earlier RTGs, where the MMRTG is a more conventional design. This similarity gives the MMRTG heritage, with an assumption of lower risk and higher trust, which is why it is positioned in the upper right corner of Fig. 2.

NASA is not the only agency that needs to predict long-term nuclear power performance in remote environments, the US Navy also has similar issues for ship and submarine power sources. Recently, the US Navy began using life-of-the-ship (LOS) reactors, which are designed to power a ship or submarine throughout its service life without refueling, which could mean 33-40 years of continuous operation [25]. Such a design change is expected to save the Navy, and thus US taxpayers, millions of dollars, but the use of LOS reactors raises safety concerns. A core design feature of LOS reactors is that the reactor is sealed, which means that the fuel and pressure vessel that contains critical coolant cannot be directly tested and inspected. The act of refueling often reveals problems that otherwise would not have been revealed because the reactor is opened [26]. Without this periodic refueling event, LOS reactors will not have additional inspections that could reveal manufacturing defects that appear in operation.

Because LOS design and testing protocols are classified, there is very little information in the public domain about such systems, but it is known that accelerated testing has been conducted at the Advanced Test Reactor (ATR) at Idaho National Laboratory (INL) for low-enriched uranium (LEU) fuel to be used in the LOS reactors. These tests

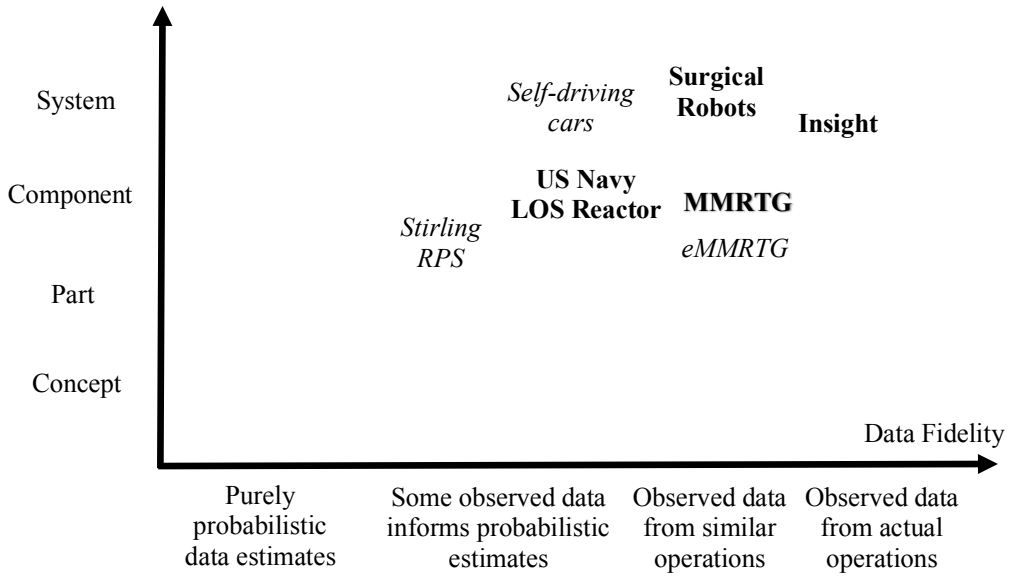


Fig. 2: Comparison of deployed (bold) and prototype (*italics*) systems published risk assessment strategies

have included irradiation of LEU fuel up to ten years to simulate fuel aging onboard a ship [27].

Based on these tests, the Navy has elected to move forward with the LOS reactor in one class of submarines with plans for additional applications [25] even though, like NASA, they cannot test the LOS systems in actual conditions for the lifetimes of the reactors, and there is no heritage for comparison. As depicted in Fig. 2, the Navy can only rely on accelerated testing of components, not even the entire system, in its risk assessment of LOS reactors. Because of the classifications around these systems, it is not entirely clear how much of the accelerated testing data relies on actual data or on more probabilistic approaches like RILT. However, it is likely that national security needs and the Navy's significant nuclear power experience and impressive safety record from the past 70 years influenced the decision to move towards a potentially riskier nuclear power source.

The last deployed system to be compared is that of surgical robots. While not a system that relies on nuclear power, it is very much a safety-critical system that can directly cause human harm. Presumably, no such technology would be deployed without a very thorough risk assessment, but as will be illustrated, the concept of heritage also occurs in medical device certification, with potentially bad outcomes.

The most common surgical robot is the da Vinci robotic surgical assistant, manufactured by Intuitive Surgical, which allows surgeons to remotely manipulate very small surgical tools through teleoperation. The da Vinci is considered a medical device and was the first of its kind, requiring FDA certification. The approval was granted in 2000 through the 510(K) process [28], which is an expedited certification process similar to NASA's concept of heritage. Medical devices are held to a lesser regulatory standard if they can show equivalence to an existing technology. Intuitive claimed equivalence to existing laparoscopic surgical technologies, which saved significant costs and development time.

Unfortunately, the decision to certify da Vinci based on its equivalence to laparoscopic surgery, in retrospect, may have been hasty. In 2014, Intuitive set aside $67 million dollars for settlements for over 3000 product liability claims with at least 52 pending lawsuits across 22 states [29]. Healthcare experts are calling for more safeguards in such systems [30], and the FDA even recently admitted problems with robotic surgery, and issued a safety communications warning doctors and patients about using such devices for cancer-related surgeries [31].

The FDA likely failed to accurately assess the risk of this new technology, the first of its kind. The FDA decided existing data for a similar system (laparoscopic surgery) would suffice in its risk assessment of robotic surgery, but in doing so, failed to understand that the two actually have very little in common. Thus, there was an illusion of equivalence, and this case demonstrates what can go wrong with an overreliance on heritage. The FDA could learn from NASA how to combine PRA approaches with existing data from similar systems to develop more comprehensive risk assessments. Correspondingly, all agencies that subscribe to an equivalence/heritage mantra (including NASA, the Department of Defense and the FAA) should be especially careful to accurately determine actual equivalence between parts, components, and systems in risk calculations.

B. Proposed Prototype Systems

The previous case studies highlight deployed technologies that were assessed, for better or worse, to have acceptable risk levels for deployment. Against this backdrop, it is instructive to examine prototype technologies that have not yet been deployed, but are in some stage of development and testing, so risk assessments are underway.

Recall from the earlier discussion that Stirling-based convertor technologies are elements of dynamic radioisotope power systems with no system flight heritage to claim. The major difference between RTG systems and DRPS with Stirling convertors is a moving piston. There have been various efforts attempting to mature this technology [24], and as discussed previously, NASA incentivized proposal teams with DRPS technology in the past. However, to date no DRPS system has flown in space and there are no planned missions, although the technology is still in development, which is why it is listed on Fig. 2 as a prototype technology.

NASA still very much needs more efficient nuclear space power for long duration space missions, so to this end, they are working with the Department of Energy (DOE) to develop the eMMRTG, a next-generation enhanced MMRTG which promises improvements over the MMRTG, including increased power across the system's lifecycle as well as an extended life from 14 to 17 years [32]. Since the technology is still under development and has not yet flown, the risk assessment is not complete. What will be interesting to follow is whether some element of heritage will be claimed for this system. The thermocouples in the eMMRTG are going to be replaced with new skutterudite (SKD)-based couples, and thus a new part will be added without heritage [33]. As a result, the eMMRTG will be on par with Stirling-based RPS systems in that the majority of the system has heritage, save for one critical component. It remains to be seen how the probabilistic risk assessment for the new SKD-based eMMRTG will be conducted and if there will be important lessons learned for Stirling-based technologies in terms of satisfactory risk assessments.

The final futuristic comparison that sheds new light on issues of risk assessment for emerging technologies is that of driverless cars. While also not a nuclear power technology, they are a safety-critical technology, with over a $100 billion dollars invested [34], and their use of probabilistic data in risk assessments is significant. Moreover, the prevalence of software in such systems is high, which are issues NASA is increasingly facing for future missions. Driverless cars are in development today by both traditional car manufacturers as well as on-demand transportation service providers like Uber and Waymo. The technology that enables cars to drive themselves includes GPS technology with detailed on-board mapping for navigation, and a combination of sensors that provide a world model for actuation decisions, which can include millimeter wave radar, camera vision, a laser detection and ranging (aka, LIDAR) [35, 36]. At the heart of decision making in these cars is a significant reliance on machine learning, often called deep learning, which makes probabilistic associations between observed and expected behaviors to determine

car behaviors. Such cars are still in various stages of development and after a pedestrian was killed by an Uber driverless car undergoing testing, there have been increasing calls for improved testing and more regulation [37].

What constitutes "good enough" testing to begin driverless cars operations is still a question of intense debate. RAND has said that such cars need to drive at least 275 million miles fatality free to be considered on par with human drivers [38]. In 2018 Waymo drove the most miles of any company, at 1.2 million [39], far short of the RAND estimates, so there has been a growing clamor in the automotive world to allow for substitutions of simulated data to demonstrate vehicle safety, e.g., [40, 41]. Driverless car companies are taking the position that the best risk assessment technique available to demonstrate to the public that the cars are safe is through simulation since it is not possible to drive the 275 million miles to gain statistical evidence in a reasonable amount of time.

One way to approach this conundrum is to develop simulations that incorporate physics-based models of vehicle dynamics so that a virtual car respects the laws of physics, such as how a car would depart a virtual road if it took a corner too fast. Then, individual sensor capabilities must be simulated including the physical dynamics of each sensor (e.g., how a LIDAR sensor responds in rain) as well as how this information would be transmitted to the deep learning algorithms that then make choices for how the car should behave. These simulations must also include representations of other drivers, including their abilities to perceive various events and then act accordingly.

Such integrated simulations that faithfully represent vehicle dynamics, sensor capabilities, and driver behavioral models are extremely complex and very expensive. No company to date has publicized any data to indicate how valid such simulations are. In the case of physics-based worlds like power generation, measured data from actual operations can be compared against computer-based simulations at very low levels of detail. While some actual data in driverless car simulations can be compared at the component level such as GPS capabilities, it is extremely difficult to validate a driverless car simulation because of the inability to faithfully represent or include every variable that leads to a critical event. In these systems, the inherent dynamics are so complex that exhaustive verification of any validation methods is beyond any set of tools available today [42, 43].

Other researchers have taken the view that more abstract simulations that embed probabilistic reasoning and risk-based approaches for holistic system safety estimates can help to determine problem areas, especially for low-probability edge cases. Such approaches resemble PRA techniques used by NASA. However, these researchers are quite clear that such simulations should be used for initial testing for problem identification, which should be followed by real-world road-based testing [43]. Still others claim that a reductionist approach to testing is the best path forward such that a form of accelerated testing called "importance sampling" can predict statistically how an automated vehicle would perform in everyday driving situations. These researchers claim that "just 1,000 miles of testing can yield the equivalent of 300,000 to 100 million miles of real-world driving [44]."

Despite the abundance of simulations proposed in the autonomous vehicle community, the lack of regulatory oversight [20] means that there is no consensus on what constitutes a "good enough" set of tests to demonstrate safety. This is particularly critical since deep learning approaches embedded in such systems have been criticized as deeply flawed [45], with an inability for human designers to effectively interpret results that likely contain bias and data overfitting [46].

While many simulation approaches used in driverless car testing bear resemblance to NASA's use of accelerated testing, the primary and critical difference is that in NASA's space power testing, hardware is the primary focus of such testing while in autonomous vehicles, software and the use of machine learning algorithms is the primary test concern. There are few, if any, parallels in terms of any heritage-like sense of confidence for driverless cars, which have not been fielded in the military, a typical proving ground for cutting-edge technology like drones. Despite the lack of consensus about driverless car testing and no existence of heritage, at least three states feel that the technology is ready, with California, Arizona, and Florida authorizing driverless car companies to begin commercial operations, although none have yet to do so.

This case is relevant to the issues surrounding risk assessment of nuclear space power because it demonstrates that there are much more questionable probabilistic-based risk assessment techniques applied to safety-critical systems with significantly less hard data for comparison than those used by NASA and affiliated agencies like DOE. It begs the questions as to why some safety-critical systems, like driverless cars and surgical robots, can be authorized to deploy with highly questionable risk assessments, while others with significant potential to advance science, like the Stirling-based convertors, struggle to convince relevant agencies that risk is acceptable? The next section will contrast and compare these industries in terms of risk assessment and provide recommendations for what could be done in the future to improve the risk assessment process.

V. Discussion

When looking across the technologies highlighted in the previous case studies, there is a complex relationship between lower-level engineers developing new or improved technologies, the agencies that oversee these projects, and the stakeholders who use and are affected by this technology. Figure 3 illustrates the conceptual ideal relationship between these entities. In a perfect world, technology developers generate a product that meets the desires of customers with minimal negative impact on peripheral stakeholders, mediated by an oversight agency that balances risk through certification and/or regulation to achieve safe use of the technology.

Take the Navy LOS case, in which the developers (the Naval Nuclear Propulsion Program, a joint effort between the DOE and Department of the Navy) have produced a derivative nuclear power technology that reduces costs without sacrificing mission performance, which is what the operational Navy desires. While this decision has not made all stakeholders happy [26], the oversight agency (National Nuclear Security Administration) has determined that the risk is low enough, likely due to the Navy's perfect safety record [25] and their undersea operational environment.

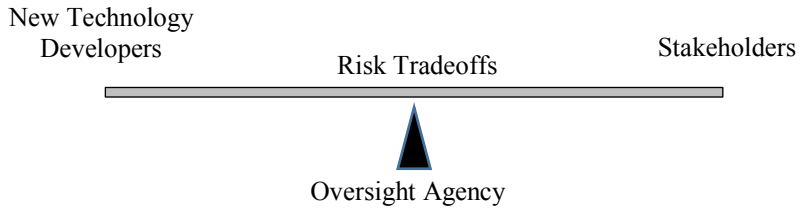
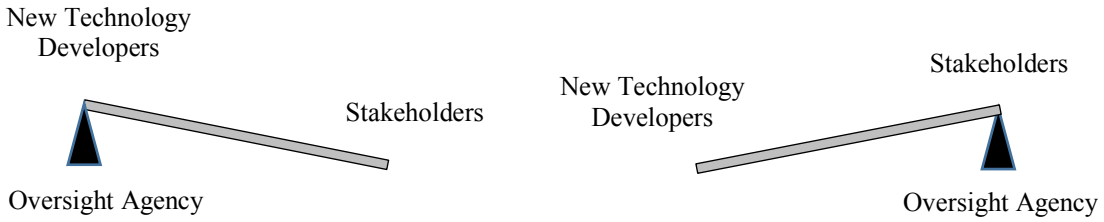


Fig. 3: Oversight agencies balance risk tradeoffs between technology developers and primary stakeholders and in a perfect world, equilibrium is achieved.

Figure 4 demonstrates what can happen when an oversight agency aligns to either side of the fulcrum, where on one side (a) technology developers are put at an advantage, and on the other side (b), they are at a disadvantage. As will be discussed more in the following sections, the recent Boeing crashes and robotics surgery systems are representative of the (a) case, while the Stirling-based convertors fall into the (b) case.

A. Oversight agency alignment with technology developers

While too much regulation can have the unintended effect of stifling innovation [23], when an oversight agency favors technology developers, as seen in Figure 4(a), then stakeholders affected by the technology can be at a disadvantage. The FAA's laxness in oversight of Boeing's 737 MAX certification has been raised as a significant



(a) Oversight agency alignment with a technology developer discounts the needs of stakeholders

(b) Oversight agency alignment with stakeholders discounts the abilities of technology developers

Fig. 4: Ramifications of misalignment of oversight agencies in new technology development.

contributor to two fatal crashes in Indonesia and in Ethiopia [47, 48]. In this particular example, the FAA delegated certification authority to Boeing, in effect allowing Boeing to self-certify their own aircraft. As depicted in Figure 4(a), this clearly puts passengers' interests at a disadvantage, and can be seen as a risk-seeking approach.

Another practice that can cause the risk balance to shift in favor of technology developers is allowing expedited certification processes to occur based on equivalence, as in the case of the FDA 510(k) process for the FDA in robotic surgery. While such practices are meant to reduce the regulatory burden and promote innovation, when new technologies are considered equivalent to older ones without a principled analysis of the potential differences, stakeholders in the form of patients suffer. Indeed, Boeing streamlined its 737 MAX certification process with claims of equivalence to the older 737 without new computer-based augmentation systems. Both the robotic surgery and

Boeing 737 MAX cases highlight that claims of equivalence between systems where one system has an entirely new software element deserves much more scrutiny.

In terms of NASA applications, when considering the role of oversight for space technologies, Congress oversees NASA's budget and the NASA Office of the Inspector General oversees NASA's program management [49], although recently the House Science, Space and Technology Committee has said it would like to have more oversight of various NASA projects [50]. There is no regulatory agency, per se, that NASA must directly answer to, but because its missions often relate to other agencies like the Nuclear Regulatory Agency and the FAA, NASA is indirectly affected by their regulations. Because NASA faces different pressures than product development companies, alignment of NASA-related oversight groups with risk-seeking technology developers tends not be the problem, but rather the opposite, discussed in the next section.

B. Oversight agency alignment with stakeholders

While NASA must ultimately answer to Congress and the public, lower-level oversight of specific missions typically falls to the Directorate sponsoring a mission, with the Inspector General's office providing higher-level oversight. Thus, for any proposed mission, the relevant Mission Directorate provides immediate oversight, often through expert-based committees. While the ultimate customer of such missions is the greater science community, the immediate stakeholders include all the centers and agencies involved, as well as the public-facing NASA headquarters.

NASA has faced continued budget pressure over the years and a fiscally-constrained high-visibility environment puts pressure on NASA to not have any incidents, which can lead to an aversion to risk. The increasing need for certainty in mission success, and the complex bureaucratic layers of NASA management can cause oversight committees to often take the most conservative path. The concept of heritage is a key element of this conservative, risk-averse path and over time, it is likely that the notion of heritage has become even more rooted. While heritage is supposed to be just one of many considerations in the selection of new mission concepts, it has become the Catch-22 for DRPS in that if such systems never fly, they can never gain heritage, which is a strong predictor for flying.

As illustrated in Figure 4(b), when oversight agencies shift the balance of risk tradeoffs towards stakeholders with heritage, and away from technology developers, the results can disfavor innovation and promote the status quo. Thus, it is a significant uphill battle for a technology like DRPS with no spaceflight heritage and a reliance on probabilistic testing to be accepted as a legitimate power supply option. This seems to be the case for the Stirling-based DRPS technologies, which continue to be passed over or replaced for proposed missions [24]. The only way for new DRPS technologies to be used in space missions, or for any newly developed NASA technology to achieve some operational use, is for the oversight agency to gravitate towards the balance point of proactively trading risk instead of the reactionary approach of minimizing risk.

The next logical questions would then be, "What influences an oversight agency to move either left or right?" The case studies presented earlier suggest that familiarity between technology developers and oversight agencies (aka regulatory capture [51]), and technical complexity beyond the capabilities of the oversight agency are significant influences. In addition, public opinion, budget and schedule pressures, the political party in power, agency culture, and media attention also play a role in how much risk is traded by an oversight agency. As evidenced in the case studies, it is paramount that oversight agencies and groups are aware of this potential bias so as to mitigate it.

VI. Conclusion

With the desire to develop a more comprehensive space exploration strategy, new advances in materials, control technologies, and power systems will require similar advances in testing and risk assessment, especially for long-duration missions. To this end, a relatively new risk assessment approach, RILT is a hybrid risk assessment methodology developed to assess dynamic RPS technologies critical for future spaceflight. It relies on accelerated testing and expert judgement for selecting testing and evaluation parameters. RILT is designed to promote risk-informed decisions while efficiently managing resource-intensive testing. It is not a substantially different approach than other NASA PRA-based risk assessment strategies, and similar strategies are used in many other domains. Despite this similarity, DRPS systems that use RILT as a risk assessment tool do not appear to be accepted in the space power community as a legitimate option.

This effort has shown that other fielded technologies have similar testing pedigrees when compared to RILT so this suggests that there are many additional factors that influence whether a technology is perceived as safe enough to further develop for deployment. This disparity can be attributed to many factors, one of which could be the lack of exposure to such a method, and to mitigate this factor, more work could be done to publish additional results from this approach, both internally and externally through professional societies.

Another likely significant factor is whether an oversight agency examining RILT is risk seeking or risk averse. Oversight groups that are risk averse tend to rely more on concepts of heritage for technology development, which ultimately results in incremental, evolutionary technologies instead of revolutionary ones, in effect stifling innovation. On the other hand, oversight agencies that are risk seeking tend to gravitate toward a concept of equivalence, which allows new technologies to be fielded even if they only share a small percentage of similarity with an older system that has been in use for some time.

It is important to note that there is a fine line between heritage and equivalence. It would be very easy for an engineer to make an argument that only one small part of a new system is changed compared to the old, and thus claim heritage, but in fact, is really claiming equivalence. So, what is actually a risk-seeking behavior is masked as a risk-averse choice. NASA and other safety-critical technology developers and regulators need to be particularly watchful for the subtle change from heritage to equivalence, especially in terms of software changes, as recent history for the Boeing 737 MAX has demonstrated how dangerous this can be.

NASA is looking towards a future of long-duration space missions where they will need next-generation nuclear space power systems that provide higher fuel efficiencies. This move will require NASA to move beyond its reliance on solar power and static radioisotope thermoelectric generators. However, because of NASA's strong reliance on heritage systems and difficulties in coming to consensus on acceptable testing data, technologies like the DRPS will continue to struggle to gain endorsement, and ultimately scientific discoveries will suffer.

Moving forward, more work is needed to determine how to best support oversight agencies both in NASA and across the government to ensure they understand their risk trade space in order to make true risk-informed decisions. For the future of long-duration space missions with increased mission complexity, developing new ways of thinking about risk assessment and mitigation is critical, as well as ensuring oversight agencies understand their alignment tendencies in the risk assessment process.

Acknowledgments

This research was sponsored by NASA, with assistance from JHU-APL and JPL. Esko Brummel provided some background research.

References

1. GAO, *Technology Readiness Assessment Guide*, US Government Accountability Office, Editor. 2016: Washington DC
2. Rose, J.R. *Risk management at JPL-practices and promises*. in *IEEE Aerospace Conference*. 2002. Big Sky, MT: IEEE.
3. Dezfuli, H., et al., *NASA Risk Management Handbook*, Office of Safety and Mission Assurance, Editor. 2011, NASA: Washington DC.
4. NASA, *Standard for Models and Simulations*, N.T. Standard, Editor. 2016, NASA: Washington DC.
5. Stamatelatos, M. and H. Dezfuli, *Probabilistic Risk Assessment Procedures Guide for NASA Managers and Practitioners* 2011, NASA: Washington DC.
6. Wreathall, J. and C. Nemeth, *Assessing risk: the role of probabilistic risk assessment (PRA) in patient safety improvement.* BMJ Quality & Safety 2004. 13(3): p. 206-212.
7. American Nuclear Society and Institute of Electrical and Electronics Engineers, *PRA procedures guide: A guide to the performance of probabilistic risk assessments for nuclear power plants* 1983, US Nuclear Regulatory Commission: Washington DC.
8. Ndu, O.K. and C. Smith. *Test Modeling Framework for Uncertainty Characterization and Life Estimation*. in *Reliability and Maintainability Symposium*. 2018. Reno, NV: IEEE.
9. Tversky, A. and D. Kahneman, *Judgment under Uncertainty: Heuristics and Biases.* Science, 1974. 185(4157): p. 1124-1131.
10. March, J.G. and Z. Shapira, *Managerial perspectives on risk and risk taking.* Management Science, 1987. 33(11): p. 1404-1418.
11. Koop, G.J. and J.G. Johnson, *The effect of multiple reference points in risky decision making.* Organizational Behavior and Human Decision Processes, 2012. 25(1): p. 49-62.
12. Dillon, R.L., et al. *Valuing Rigor in the Risk Management Process*. in *IEEE Aerospace*. 2019. Big Sky, MT: IEEE.
13. Under Secretary of Defense for ATL, *Highly Accelerated Life Testing and Highly Accelerated Stress Screening Methodology*, Department of Defense, Editor. 2016: Washington DC.
14. White, C.C. and D.L. Hunston, *Accelerated Testing: Challenges and Opportunities*. 2017, Gaithersburg, MD: NIST.
15. Wiksten, D. and J. Swanson, *Accelerated Life Testing of Spacecraft Subsystems*, Jet Propulsion Laboratory, Editor. 1972, NASA: Pasadena, CA.
16. Murray, S.F., H. Heshmat, and R. Fusaro, *Accelerated Testing of Space Mechanisms*. 1995, Mechanical Technology Incorporated: Latham, NewYork.
17. Ghaffarian, R. *SIP Qualification and PRA Approaches*. in *Surface Mount Technology Association International* 2002. Boston.
18. Meshkat, L. *Probabilistic risk assessment for decision making during spacecraft operations*. in *Annual Reliability and Maintainability Symposium*. 2009. Fort Worth, TX.

19. Hirshorn, S.R., *NASA Systems Engineering Handbook Revision 2*, Aeronautics Research Mission Directorate, Editor. 2017, NASA: Washington DC.
20. Cummings, M.L. and D. Britton, *Regulating Safety-Critical Autonomous Systems: Past, Present, and Future Perspectives*, in *The Psychology of Interacting with Robots*, R. Pak, E.d. Visser, and E. Rovira, Editors. 2019.
21. National Research Council, *Priorities in Space Science Enabled by Nuclear Power and Propulsion*. 2006, The National Academies Press: Washington, DC.
22. Petroski, H., *To Engineer is Human*. 1992, New York: Vintage Books. 251.
23. Bennett, D., et al., *Annual Report of the Government Chief Scientific Adviser 2014. Innovation: Managing Risk, Not Avoiding It. Evidence and Case Studies.* 2014, UK Government Office of Science: London.
24. Brummel, E.S., et al. *Identifying and Mitigating Barriers to the Adoption of Dynamic Radioisotope Power System for Space Flight*. in *IEEE Aero Conf*. 2019. Big Sky, MT: IEEE.
25. Moore, G.M., ed. *The 6 Percent Solution: LEU Fueled Reactors and Life-of-Ship Reactors for the US and UK Navies*. Institute for International Science & Technology Policy Occasional Papers Series: Reducing Risks from Naval Nuclear Fuel, ed. P. Lobner, et al. 2018, The George Washington University: Washington DC.
26. Moore, G., *Life-of-Ship Reactors and Accelerated Testing on Naval Propulsion Fuels and Reactors*. 2017, Federation of American Scientists: Washington DC.
27. National Nuclear Security Administration, *Conceptual Research and Development Plan for Low-Enriched Uranium Navsal Fuel*. 2016, US Department of Energy: Washington, DC.
28. Jones, U. *FDA Clears Robotic Surgical System*. 2000 [cited 2018 July 17]; Available from: https://www.meddeviceonline.com/doc/fda-clears-robotic-surgical-system-0001.
29. Baron, E., *Robot-surgery firm from Sunnyvale facing lawsuits, reports of death and injury* in *The Mercury News*. 2017, Bay Area News Group: San Jose.
30. Sheetz, K.H. and J.B. Dimick, *Is It Time for Safeguards in the Adoption of Robotic Surgery?* JAMA, 2019.
31. US Food and Drug Administration *Caution when using robotically-assisted surgical devices in women's health including mastectomy and other cancer-related surgeries*. FDA safety communication, 2019.
32. Woerner, D., *A Progress Report on the eMMRTG.* Journal of Electronic Materials, 2016. 45(3): p. 1278–1283
33. Holgate, T.C., et al., *Increasing the Efficiency of the Multi-mission Radioisotope Thermoelectric Generator.* Journal of Electronic Materials, 2015. 44(6): p. 1814–1821.
34. Leasing Options. *Over $100 Billion Invested In Driverless Technology*. 2019 [cited 2019; Available from: https://www.leasingoptions.co.uk/driverless-cars/index.html.
35. Hicks, M. and M. Fitzsimmons. *Driverless cars explained: everything you need to know about the futuristic tech*. Tech RAdar 2018 [cited 2019 May 11]; Available from: https://www.techradar.com/news/driverless-cars-explained.
36. Rychel, A. *The different sensor technologies explained.* 2017 [cited 2019 May 11]; Available from: https://www.2025ad.com/latest/driverless-cars-infographic-sensors/.
37. Laris, M., *Fatal Uber crash spurs debate about regulation of driverless vehicles*, in *Washington Post*. 2018, Nash Holdings: Washington DC.
38. Kalra, N. and S.M. Paddock., *Driving to Safety: How Many Miles of Driving Would It Take to Demonstrate Autonomous Vehicle Reliability?* 2016, RAND: Santa Monica, CA.
39. Madrigal, A.C., *Waymo's Robots Drove More Miles Than Everyone Else Combined*, in *The Atlantic*. 2019: New York.
40. Dent, S. *Toyota will be first to use NVIDIA's self-driving simulator*. 2019 [cited 2019 May 11]; Available from: https://www.engadget.com/2019/03/18/toyota-nvidia-drive-constellation-simulator/.
41. Madrigal, A.C. *Inside Waymo's Secret World for Training Self-Driving Cars*. 2017 [cited 2019 July 11]; Available from: https://www.theatlantic.com/technology/archive/2017/08/inside-waymos-secret-testing-and-simulation-facilities/537648/.
42. Henzinger, T.A., et al., *What's decidable about hybrid automata.* Journal of Computer System Sciences, 1998. 57(1): p. 94–124.
43. O'Kelly, M., et al. *Scalable End-to-End Autonomous VehicleTesting via Rare-event Simulation.* in *32nd Conference on Neural Information Processing Systems*. 2018. Montreal, CA.
44. Zhao, D. and H. Peng, *From the Lab to the Street: Solving the Challenge of Accelerating Automated Vehicle Testing.* 2017, MCity: Ann Arbor, MI.
45. Marcus, G., *Deep Learning: A Critical Appraisal.* arXiv:1801.00631, 2018.
46. Koopman, P. and M. Wagner, *Challenges in Autonomous Vehicle Testing and Validation*, in *Society of Automotive Engineers Wrold Congress*. 2016: Detroit.
47. Cassidy, J., *More questions than answers about Boeing, the 737 MAX, and the F.A.A.*, in *The New Yorker*. 2019, Conde Nast: New York, NY.
48. Gates, D., *Flawed analysis, failed oversight: How Boeing, FAA certified the suspect 737 MAX flight control system*, in *Seattle Times* 2019, Blethen Corp.: Seattle.
49. OIG. *NASA IG*. 2019; Available from: http://oig.nasa.gov.
50. Klimas, J. *New House Science chair seeks more oversight of NASA projects*. 2019 [cited 2019 May 26]; Available from: https://www.politico.com/story/2019/01/16/congress-space-house-nasa-eddie-bernice-johnson-1101691.
51. Stigler, G.J., *The Theory of Economic Regulation.* The Bell Journal of Economics and Management Science, 1971. 2: p. 3-21.