SAFETY II WHITE PAPER: ASSESSING RESILIENCE AND HUMAN VARIABILITY WITHIN AVIATION SAFETY

GARETH COVILLE
DHARM GURUSWAMY
BRIAN HILBURN, PHD
CHUCK HUBER
HOUDA KERKOUB
GENE LIN, PHD
DAVID MCKENNEY
GREG SIZEMORE

NOVEMBER 2024

MITRE | SOLVING PROBLEMS FOR A SAFER WORLD®

NEW COMPLEMENTARY APPROACHES ARE NEEDED TO IMPROVE UPON EXISTING FAA SAFETY ACCOMPLISHMENTS.

In 2008, the Federal Aviation Administration (FAA) set a goal to reduce commercial aviation fatalities per 100 million people onboard by 50% by 2025 [1], which has been successfully achieved. This goal was after the FAA had successfully worked with the Commercial Aviation Safety Team (CAST) to "reduce the fatality risk for commercial aviation in the United States by 83 percent from 1998 to 2008" [2]. This success over the last quarter century was due in part to applying traditional data-driven safety approaches to proactively identify combinations of precursors that could lead to an accident and then effectively mitigating them. Despite these successes, traditional data-driven safety approaches of addressing precursors to accidents are becoming more challenging as our nation's aerospace system gets ever more complex.

SAFETY II: LEARNING FROM "WHAT IS GOING RIGHT"

Safety II offers a complementary approach to the FAA's traditional safety management approaches and can be summarized as "mov[Ing] from ensuring that as few things as possible go wrong to ensuring that as many things as possible go right" [3]. Safety II acknowledges that humans bring variability to complex sociotechnical systems, and as a result, precursors will always exist. Our nation's aerospace system is an example of a complex sociotechnical system that consists of both the daily operations of the National Airspace System (NAS), and the processes employed to design, build, certify, and maintain aerospace products. The FAA and other aviation stakeholders can continually improve the safety of the aerospace system by learning from all operations (i.e., asking what is working well in addition to focusing on what went wrong) and understanding how human variability is contributing to safety.

Three principles that Safety II shares with other systems-safety perspectives:

1. Human variation in complex systems is inevitable. This variability can, yet seldom does, result in negative consequences; its presence is mostly positive and arguably essential.
2. Human variation should not be considered the root cause of failed outcomes but rather symptoms of poor system design.
3. We can improve safety by improving what is going wrong and learning from what is going right (i.e., learning from all operations).

By applying these principles, the FAA can help its system designers and safety experts to better understand how human variability helps our nation's aerospace system to remain safe despite its complexity.

However, a literature review conducted by The MITRE Corporation (MITRE) [4] found that there are many theoretical views on Safety II, but few operational applications exist that have manifested in immediate, positive impact for either the government or industry.

SAFETY II METHODS FOR THE FAA

A study of the Safety II literature, and reflection on how safety is practiced in aviation and in other fields, has led us to three methods for improving safety practice at the FAA. These methods are versatile enough to be applied across all safety offices within the FAA. The methods are described as follows:

1. *Barrier Frameworks.* When the FAA enumerates existing safety barriers and their relationships to each other, it becomes possible to categorize historical events based on which barriers worked as expected and which did not; and furthermore, to ask structured questions around those courses of events.

Expected impact: By formalizing the understanding of how existing risk controls work together, the barrier framework provides an approach for describing safety systems – a required step envisioned by FAA's *Order 8040.4C - Safety Risk Management Policy* [5]. Also, by understanding barriers and their interactions, we can endeavor to build similar structures of different barriers for new entrants and operations, such as those for Advanced Air Mobility or Commercial Space operations. Since there will not be data on operational safety for some time, having approaches to build and understand safety is paramount for their successful introduction.

2. Resilience Assessments using the MITRE TRUSTS Framework [6]***.*** The TRUSTS Framework provides a standardized set of eighteen questions used to evaluate system resilience based on a specific event from the perspective of individual agents within the system as well as the overall system. The resilience insights gained from individual events can then be scaled using a FAA-specific Large Language Model (LLM) populated with thousands of safety reports to determine if these behaviors are regularly observed in the system.

Expected impact: This method enables the FAA, airlines and other partners to better articulate resilient best practices that agents take to ensure a safe outcome, and then ensure that these best practices are reinforced during training so that other agents will be empowered to take the same action when faced with a similar situation.

3. Gap Analyses. A primary pillar within the Safety II literature is the concept of Work-As-Imagined (WAI) versus Work-As-Done (WAD) [3]. Specifically, WAI details how a complex sociotechnical system, like the nation's aerospace system, is expected to operate from a system designers' perspective. However, due to a variety of pressures put on agents and situational variability within the system, agents may modify their tasks based on a tradeoff between being efficient, being thorough, or somewhere in between. As a result, the WAD that an agent undertakes to complete a task could vary significantly from WAI.

Using a LLM populated with FAA guidance documentation (i.e., WAI) [7] and voluntary safety reports (i.e., WAD) allows the FAA to gain insights into (1) what FAA guidance documentation is available for agents when barriers degrade, and (2) what agents document in their safety reports when these barriers degrade.

Expected impact: This method is important because it allows the FAA to identify and mitigate gaps between WAI and WAD before an accident is realized. It also identifies common resilience themes that can be promoted as safety best practices.

RUNWAY INCURSION USE CASE FINDINGS

We applied these three Safety II methods to Runway Incursions (RIs) as a use case. Specifically, we conducted a Resilience Assessment on the RI that occurred at Austin-Bergstrom International Airport on February 4, 2023; applied Barrier Frameworks to four recent RIs; and studied 1,700 reports of RI incidents between October 2018 to February 2024 for a Gap Analysis. Our findings were:

- Agents exhibit resilience when avoiding RIs, but this behavior is often taken for granted and goes unnoticed. For example, our Resilience Assessment highlighted that the shared safety responsibility and quick actions shown by the flight crew in the arriving aircraft resulted in a narrow escape from a potentially catastrophic accident. By using the three methods above, the FAA, airlines, and other operators, could highlight observed resilient behaviors and incorporate them into their RI training.
- The existing network of barriers for preventing RIs provides redundancy if any single barrier fails. The three methods we recommend identified examples to demonstrate how these barriers create a resilient runway environment that minimizes fatal RIs. For example, we found that by using the Barrier Framework method to timeline the barriers involved in four RI incidents, it was possible to quickly identify which barriers failed and which succeeded. This analysis highlighted how the system of barriers developed by the FAA prevented the RIs from escalating to catastrophic accidents, even though individual barriers failed during each incident.
- Human variability is a contributing factor to causing and resolving most RIs. FAA orders for NAS-users should better account for human variability. The three methods identified above identified examples of how this variability permeates the NAS. For example, our Gap Analysis showed that human variability is often a contributor to causing a RI, and this same human variability is a primary reason why RIs seldom escalate to a fatal accident[1].
- Pilot expertise such as airmanship and effective teamwork are crucial to maintaining safe operations and until now, they have been challenging to document in the operational environment.

[1] The last fatal RI that occurred in the NAS was the Comair Flight 5191 accident in 2006.

FAA APPLICATIONS OF THE SAFETY II METHODS

Below are examples of how the three Safety II methods could be applied in AVS:

1. Performance-based regulations and policy. The three Safety II methods outlined above can aid the FAA in developing performance-based regulations. Regulations act as risk controls in the aviation system and are developed using the Safety Risk Management (SRM) process. The first step in the SRM process is to understand the system context. Applying the Barrier Framework will help the FAA better understand where risk controls should be applied. The Resilience Assessment and Gap Analysis as applied to the Barrier Framework will guide the rulemaking team to promote resilient behavior in the regulations thus making them more performance based. These concepts can be equally applied to development of policy and guidance.

2. Oversight and Surveillance. In the future, the FAA should approach oversight of product and service providers through a Safety II lens. The FAA (in service units like the Aircraft Certification Service (AIR)) already requires original equipment manufacturers, as part of its product assurance, to specify what bad outcomes they are guarding against and the risk controls for those outcomes. Furthermore, industry should be reporting when they have discovered a better way of enhancing safety, and the FAA should support enhancing industry's systems based on these discoveries. In other words, the FAA should purposely look for where adaptive behavior is creating higher levels of safety and develop mechanisms to collect the associated data and feed it back throughout the safety lifecycle. The Resilience Assessment and Gap Analysis could focus FAA resources towards finding these examples that promote high performance adaptive behaviors.

In addition, performing oversight of product and service providers is intended to create continual improvement in safety, mainly by identifying shortcomings in a company's safety system. The find and fix mentality is necessary, but not optimum for increasing aviation safety. What goes right occurs more frequently than what goes wrong, yet the finding and fixing of what goes wrong are where resources are focused. Safety could be more rapidly improved if positive behaviors that improved safety were detected, shared and leveraged throughout the aerospace system. The challenge is that these positive behaviors may not be considered "compliant" behavior.

The three Safety II methods can help determine if WAD, while maybe not compliant, achieves the intent of the WAI and upholds acceptable levels of safety. By applying these methods, the FAA could determine that some non-compliant behaviors found during oversight might result in higher levels of safety. If this adaptive behavior is found to be safer, then it should be promoted throughout the aerospace system.

3. Certification. Current regulations struggle to keep up as the aviation industry continues its rapid pace of innovation. Applying safety requirements from the standpoint of accentuating what goes right to manage safety instead of just considering what goes wrong, can help the FAA make decisions regarding acceptable levels of safety. Systems that promote continual improvement to doing things well can reduce the likelihood of things going wrong. The Resilience Assessment and Gap Analysis can be used to promote both system and personnel adaptive behavior in the management of safety.

NEXT STEPS: FAA SHOULD ADOPT THE SAFETY II METHODS

The next steps that MITRE recommends that the FAA should pursue are:

- *Adopt the three Safety II methods across AVS to identify best practices and understand what is going right* (i.e., analysis of resilience and effective barriers), in addition to focusing on what is going wrong (i.e., analysis of incidents and accidents). Once these best practices are identified they should be incorporated into FAA guidance and training documentation.
- *Explore how the different FAA voluntary safety reporting systems can better capture what went right during an event – what barriers were in place, and what actions were taken, to de-escalate a developing hazard*. This sort of reporting would not only enhance the FAA's understanding of how barriers and resilience are applied to provide safety, but also might point towards the existing behaviors and barriers that are currently the most effective.
- *Leverage the Barrier Framework to describe relevant components of the aerospace system when conducting a safety assessment* as required in FAA Order 8040.4C.[2]
- *Apply the Resilience Assessment and Gap Analysis to support the FAA's future State Safety Program (SSP) initiatives.* For example, the FAA could perform a Resilience Assessment and Gap Analysis on a corpus of voluntary safety reports organized by incident type to create a set of best practices per incident type and then bring those positive insights to the SRM panels when they are developing new controls. Those same positive insights could also be used to support the FAA's Safety Promotion initiatives. As well, resilience analysis takeaways in terms of "what went well" can extend beyond SRM and support Safety Assurance assessments in the current environment of sparse accidents and incidents.

CONCLUSIONS

This paper has outlined three practical Safety II methods for assessing safety via:

- *Barrier Frameworks* to understand when and how barriers succeed or fail.
- *Resilience Assessments* to articulate and categorize observed resilience examples.
- *Gap Analyses*, to analyze existing rules and work practices to identify gaps between how system designers envision tasks to be completed and how agents actually complete their tasks.

As shown above, these three methods provide insights into resilience and human variability that would not be identified using traditional safety approaches. We believe that these three methods can be readily applied by the FAA to enhance the effectiveness of existing safety assurance processes across a broad range of domains including performance-based regulations, oversight and surveillance, and certification. By adopting these three Safety II methods, the FAA would have additional methods to support its SSP processes, new insights as to why the NAS is operating safely, and recommendations for how to leverage these insights to further enhance NAS safety.

[2] Required as part of the System Analysis Step on page 13 of FAA Order 8040.4C

GLOSSARY

Acronym	Definition
CAST	Commercial Aviation Safety Team
FAA	Federal Aviation Administration
LLM	Large Language Model
MITRE	The MITRE Corporation
NAS	National Airspace System
RI	Runway Incursion
SSP	State Safety Program
SRM	Safety Risk Management
WAD	Work as Done
WAI	Work as Imagined

BIBLIOGRAPHY

[1] FAA, "FAA Portfolio of Goals," July 2023. [Online]. Available: https://www.faa.gov/sites/faa.gov/files/FY23%20Portfolio%20of%20Goals%20July%202023.pdf.

[2] CAST, "Commercial Aviation Safety Team," 2022. [Online]. Available: https://www.cast-safety.org/apex/f?p=102:1. [Accessed 23 October 2024].

[3] E. Hollnagel, R. L. Wears and J. Braithwaite, "From Safety-I to Safety-II: A White Paper.," 2015.

[4] MITRE, "Moving Safety II from Theory to Practice, Enclosure of CAASD Product 4-4.3-1," McLean, VA, 2024.

[5] FAA, "Order 8040.4C - Safety Risk Management Policy," Washington DC, 2023.

[6] MITRE, "MITRE RAD (Resilience-Aware Development) & The TRUSTS Framework," 2024. [Online]. Available: https://trusts.mep.mitre.org. [Accessed 22 August 2024].

[7] E. Mangortey, "Artificial Intelligence Risk Assessment for Large Language Models, Product 4-4.1-2.," MITRE, McLean VA, 2024.

NOTICE

This work was produced for the U.S. Government under Contract 693KA8-22-C-00001 and is subject to Federal Aviation Administration Acquisition Management System Clause 3.5-13, Rights In Data-General (Oct. 2014), Alt. III and Alt. IV (Oct. 2009).

The contents of this document reflect the views of the author and The MITRE Corporation and do not necessarily reflect the views of the Federal Aviation Administration (FAA) or the Department of Transportation (DOT). Neither the FAA nor the DOT makes any warranty or guarantee, expressed or implied, concerning the content or accuracy of these views.

For further information, please contact The MITRE Corporation, Contracts Management Office, 7515 Colshire Drive, McLean, VA 22102-7539, (703) 983-6000.