Flight Test *Safety* Fact

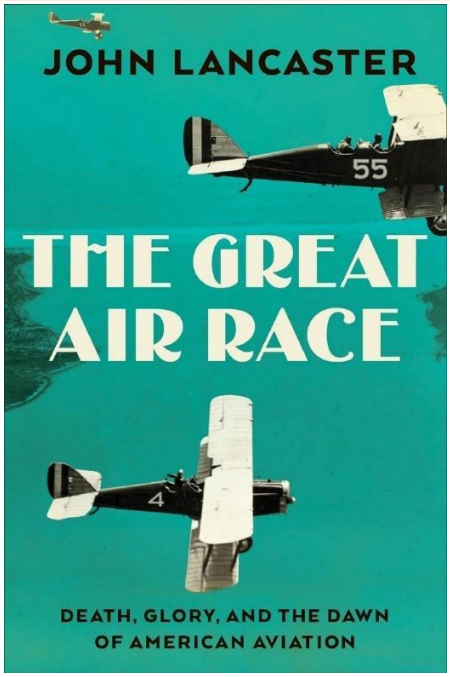
Published for the Flight Test Safety Committee

In This Issue

The Stage - a collection of miscellany relevant to all things flight test safety and setting the stage for a flood of reader letters
Letter to the Editor: "My Experience with STPA" - Brian Ostermann shares a two part letter sharing his STPA experience
Letter to the Editor: "Answers to Rhetorical Questions" - Ben Luther comes out of his "convent" to advocate for STPA
Letter to the Editor: "Pancho Weighs in" - The original STPA speaker of 2023 FTSW fame answers the editor's inquiry too
Turbo Talk - How well do you take advantage of simulators? That question appears in the Talk column and the latest podcast!

The Stage

Mark Jones Jr



I just finished reading *The Great Air Race: Glory, Tragedy, and the Dawn of American Aviation*, by John Lancaster, which claims to be "The untold, almost unbelievable, story of the daring pilots who risked their lives in an unprecedented air race in 1919—and put American aviation on the map." I thought it lived up to the bold claim.

The cover of the book is beautiful—an observation I will come back to—and the story is riveting. One might read it simply for entertainment, but it is immediately relevant to any branch of aviation safety, including flight test safety. I can also think of several other domains where the subject matter applies, including innovation and change, history of science and technology, adventure, uncertainty, and a host of other topics. For example, Advocates for UAS and eVTOL and every other buzzword and acronym would benefit from the story.

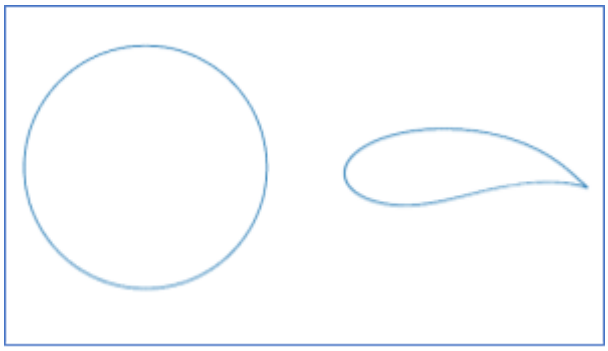
The book included several biographical sketches on Billy Mitchell, information I had not previously known or from a different perspective or both. In general, I was left with a distaste for many of the contemporary leaders of that era.

On the other hand, I have a new appreciation for the pilots of the era. They were brave and resourceful, far more so than we are today. Sometimes they were stupid, just like we still are, sometimes. Once in a while they were lucky, and sometimes luck was simply the manifestation of skill and preparation meeting opportunity.

I want to come back to the comment I made about the cover design. I think it's beautiful, and that's simply a matter of taste, of aesthetics. We don't often talk about art or "right brain" stuff but Boeing's X-32 is an amusing reminder that maybe we should. For years, I encountered this kind of "artistic enrichment" as a regular reader of "This Week in CFD" by John Chawner. The weekly column was an eclectic collection of CFD news, and it often included a photograph of artwork the author fancied, something that both appealed to his taste and reminded him of the kind of grids and geometry he encountered in his work. I enjoyed it and wondered why we don't do something similar.

I've also recently added The Aperiodical to my RSS feed reader, because it does something similar in the domain of mathematics and just so happens to feature a book cover in its most recent post. But if you go two posts back, the miscellany covered was an airfoil shape I'd never encountered. The image (below) and the math are both fascinating. Another source of visual media is the weekly General Aviation News, which does a good job of

sharing a "Picture of the Day." Sometimes the photos are stunning because of the skill of the photographer, and other times it is the subject of the photo—an airplane or a sunset or both—that catches the eye. Coincidentally, I have an entire folder on my computer for pictures of the sky, many of which include some airplane appendage in the corner of the photo but just as many that I've shot from the ground.



Credit: https://aperiodical.com/2023/08/mathematical-objects-joukowsky-aerofoil/

These things are all both fanciful and technical, and both can enrich us as professionals—even if you wondered where this random walk is going—and for those counting, there are two more stops for our traveling salesman. The first is a shotgun pattern of three topics: a flipping penny, a deck of cards, and a brown bag with (maybe) colored marbles in it.

1. Heads or tails?

You can imagine pulling a coin out of your pocket. There are two possible outcomes, heads and tails. If this is a fair coin, each outcome is equally likely. This is probability without numbers, but it's also perfectly precise. We can extend it to a fair dice. The probability of each outcome (each of the sides of the dice) is the same. They are equally likely. However, we can also say that getting an even number is more likely than tossing a one. More likely and less likely are both understood in this sense, even though it is not as precise. There are two things I want to emphasize. First, probability predicts. I hope the alliteration in the previous sentence helps you remember what probability is and what it isn't. Second, probability is a model of the world. Nothing is truly random, including the motion of molecules in a glass of water, but we can use probability as a useful model. As long as you understand that it is a model and that there is a distinct difference between probability and statistics, there's a chance you can employ it correctly.

2. Black Swans, or "What are the chances?"

Unlikely things happen all the time. Imagine that you had a deck of cards, which are shuffled fairly. Also imagine that you deal yourself the following hand.

Do you remember how to compute the probability of getting this hand of cards? I'll save us all some frustration and tell you the answer: The probability of getting this hand of cards is 3×10^{-9}. That is incredibly unlikely. Extremely unlikely things happen all the time. So maybe, just maybe, don't be surprised the next time you see a black swan (which is a bad example of unlikely things because so many of them exist). If you are following along

at home with the math, we can have a separate conversation about how I computed that probability, because it requires us to clarify our assumptions.

3. N=1
Use your imagination one more time and pretend that I just gave you a brown bag full of something that probably will turn out to be solid-colored balls. Really, though, you have no idea what's in it. Magically, you reach your hand inside the bag without seeing anything—because that's what the theory and the math book require—and you pull out a white round object. Your sample size is one. Some readers will immediately say that this is not statistically significant, but opinions may vary. I would argue that you now know definitively that the bag contained a white ball. There is no uncertainty in that fact. The sample size doesn't matter. If you put the ball back in the bag, your knowledge of the uncertainty is higher than it was before the first pick. It is not much higher, but you have learned definitive, incontrovertible information from a sample size of N=1. That's significant.

These three anecdotes are just fanciful sights that one can observe on a wandering stroll through the topic, threads that make up a rich fabric which finds its way into our profession, into safety and measurement and so many other things. But I promised one more stop on this journey, and I'll introduce this final waypoint by saying this: If you look back at history, it's usually the letters and newspaper clippings that we piece together, thus creating a mosaic, and only when we step back can we appreciate the whole picture and comprehend the object rendered. This observation sets the stage for what's next.

Thus it is only appropriate and timely that I have three letters to share with you, correspondence motivated by the last issue of the Flight Test Safety Fact. I've enjoyed the exchange of ideas and appreciated the willingness of these members of our profession to contribute their observations. I hope you engage them as well.

Letter to the Editor: "My Experience with STPA"

Brian Ostermann

Part 1
Mark,
After reading the August 2023 Flight Test Safety Fact and your STPA Deep Dive, I encourage you to explore the MIT Partnership for Systems Approaches to Safety and Security (PSASS) web page located at http://psas.scripts.mit.edu/home/ and all its resources. I also encourage you to explore the tutorials and download the handbooks.

Every year in June Nancy Leveson and John Thomas at MIT host a workshop presented both "live" and virtually; I also encourage you to sign up for it. Additionally, I highly recommend reading Nancy Leveson's book Engineering a Safer World. This book should be the system safety baseline bible for every engineer and is the basis to explaining STPA.

I applaud Mirf's and Pancho's efforts in spreading the good word with STPA in the tutorials posted on the Flight Test Safety website, but I thought it more technical than what the subject needs to be.

The other problem that no one really discusses, which often gets blindly put into many requirements, is referencing Mil-Std-882—it's old; and how many engineers truly get trained to implement PHAs, FHAs, FMEAs, FMECAS, FMETs, RCAs, FTAs, and UERAs just to name a few of the many acronyms associated with safety analysis tools out there.

Long story short, with STPA you will get more bang for your buck in the end.

Mark: *I have been to the website before, but like many FT professionals, it seems like I don't have enough time to dig deeper without having a compelling "why." Is it just a system safety tool to replace FMEA or can we use it effectively for flight test safety...is it truly a better way to do THAs?*

Part 2

All my experience with STPA has been either informal to semi-formal. I say informal because all my familiarity has been self-study and peer discussions; I say semi-formal where I have made STPA a recommendation in a number of after action or test reports where I have seen absolutely no fault mode or hazard analysis completed at all. STPA is strictly a technique in the culture right now.

Towards the end of grad school, I came across the following excerpt from Human Factors in Aviation, 2nd Edition, by Eduardo Salas and Dan Maurino, 2010: "Evaluating for system safety and predicting what can go wrong: How should the system work? What can go wrong and how should the system respond?" This triggered something that I now call Brian's technique to system safety and hazard analysis when I perform any evaluation.

Brian's Technique

The System	The Human
What is the system; know the system, learn the system. How does the system function and respond? How does the system fail? How should the system respond when the system fails?	Who is the human operator; know the operator, learn the operator. Where and in what environments does the human operator operate? How does the human operator respond to the system? What can go wrong with the human? How does the human respond when the system is perceived to fail?

But now you also have to look at the two components completely together as one system and not separately.

The Human + System
Does the operator's mental model with using the system match the same model and intent from the designer? How, where, and why are the human and system integrated? What can go wrong with the integration? Can vulnerabilities be identified? Can risk be eliminated, reduced, or accepted?

The above technique is not STPA, but rather a hybrid I figured out after the fact. The common denominator is the controller taking a top down approach. A former classmate, Shem Malmquist (Fed Ex pilot, Accident Investigator, Professor) started working with Nancy Leveson and John Thomas and started sharing some of their work. That in turn steered me to look more into STPA/STAMP/CAST.

So here is what I've learned in a nutshell:

All of our traditional system Safety analysis tools are not stand alone. More than one tool must be used to get a complete picture if a solid hazard analysis is to be completed.

FMEA, FMECA and FMET take a bottom up approach and are usually restricted to a specific widget within a system for component failure. Only recently has uFMEA ("usability" FMEA or "Use Error" FMEA) really had any exposure (Risk Management, Israelski and Muto, 2022), but it still requires multiple tools over lengthy time.

FMEA, FMECA and FMET are all driven by either 882 system safety or 516 airworthiness. Mil-Std-882 is old and outdated.
STAMP and STPA take a top down approach beginning with the controller (human or computer/widget). I can get more bang for my buck with an STPA analysis bringing in all stakeholders. Therefore, STPA can and should be used throughout the entire design and evaluation process much like the last step in any risk management process.

A paradigm shift needs to take place if STPA is to be successful; it is only a technique at this point. Much like many of the tools we use today that were created many years ago.

I have also found that very few people have been trained in completing FMEAs, FMECAs, and FMETs or aren't familiar with the history behind them. Or I have found that they take their best guess at a hazard completing the black, red, yellow, green probability table. Don't get me wrong here, they are good tools as long as those using them understand their process and their limitations. Recently I was brought onto an acquisition program to help find the root cause to a system problem. One of the first questions I asked was, "where is the FMEA for the system?" The reply I received, "we don't know, it's too far down the road to find or complete that." Are you kidding me? That could have been the answer to the traceability to the problem, or at least provided a clue.

In another example I extracted from an email reply from a customer to a report I submitted to them involved a number of recommendations:

(ME) R19. Complete an Failure Mode Effects Test (FMET) to determine the effects of single failures and combination failures during the XXXXXXX_XXXXXXX checklist initiation process.
(CUSTOMER) What is your definition of a Failure Mode Effects Test in this scenario? Common understanding: input the failure modes into system, see how system responds?
(CUSTOMER) What is the benefit of doing this test? Assure that the XXXXXX allows the aircrew to perform mitigation procedures after a failure?

As for THAs specifically, I've seen the personality of organizations drive THAs. What defines a test hazard, how a test hazard is identified, and how a test hazard might be mitigated?

Unfortunately, the most training anybody in the DoD might get in system safety or hazard analysis is their interpretation of Mil-Std-882; if they actually read it. In my case, it was grad school exposure at both FIT and UCF that completed the puzzle for me. Otherwise it was baptism by fire during my time in the military, or I relied on my operational flying risk management tools.

Brian
<Brian.Ostermann@sti-tec.com>

Letter to the Editor: "Answers to Rhetorical Questions"

Ben Luther

Mark,
Thanks for continuing to deliver a great newsletter for many years. It's great because it serves as a platform for discussion in our flight test community – you are meeting your goal. At the risk of writing a letter to the editor (how old is that!?), maybe I can contribute via the rhetorical questions you pose in the August 2023 edition. As someone now working in flight-test-adjacent fields (hoping to find a way back one day), I've taken myself to your rhetorical convent in a non-flight-test context where I research the way Flight Test manage risk with a view to using that elsewhere.

My convent is the University of Adelaide, and I don't get any kudos for anything to do with aeroplanes. They simply aren't impressed by aviation (what's wrong with them!?). But I have had positive academic review with my examination of why flight test crews are successful at managing risk. Many in the SETP and SFTE communities answered my call late last year by volunteering their practices toward my research. My thanks to them. (For the PhD candidates keeping score, 81 responses for n=49. Not bad for social science.) Flight test is one of the few (only?) industries that routinely operate across three domains of system intricacy and researching this community is enabling abstraction of generalisable principles. A key learning moment in the research was the universal reporting of mentoring and review. That practice is evidence toward my thesis that risk has domains and risk management tools are domain specific. I argue this is the answer to your question "What is the value proposition of STPA?". One-size does not fit all, and STPA is a tool that is effective in the system intricacy domain labeled Complex. It is overkill in the Clear and Complicated domains where statistical approaches are valid and cheaper options. Flight Test uses statistical tools, but they retain non-statistical approaches at their discretion. That discretion kicks in when we encounter complexity.

Using a Cynefin lens to define these domains, risk in the Clear domain is deterministic and has low latency. My hammer, my thumb, my feedback loop. The consequences are bounded by low latency that precludes third parties becoming involved. Low cost consequences (from a societal perspective) don't warrant expensive management approaches and high occurrence rates support statistical analysis. The Complicated domain remains deterministic (cause and effect correlations are apparent) but introduces a latency. There is a best way to undertake an activity and it can be known. It might not be easy, but all the options can be known in advance. Because it is deterministic we learn the best way and do that every time. This is the realm of commercial aviation and statistical approaches are valid because the system configuration is constant.

It gets interesting now that our systems have evolved to access the Complex domain. Complex systems feature emergence (recall Montes' definition in the STPA lecture) so they cannot be known in advance. Emergence precludes a top-down decomposition since the emergent functions don't reside in any one component. The systems are no longer deterministic and latency still features. (Remove the latency and you are in Chaos where the system is unstable.) Complex systems are dynamic so knowledge is perishable, and the dynamic nature invalidates statistical approaches as the system is different at each observation. In other words, n=1 doesn't offer a statistical model. No-one should be using a 2D hazard matrix on a complex hazard. For that, invest in learning STPA.

To your question of "how robust is the analysis process?" I suggest the question is rooted in a systems perspective of decomposed components, rather than a functional perspective. The power of model-based system engineering is in allowing different perspectives. STPA adopts a functional abstraction, and Montes' diagram of the system inside the environment is relevant. The system designer decides what functions are in the system boundary and this correlates with functions that are controlled. All that aren't controlled are the environment. Leveson's observation of functional failure being independent of component failure is pivotal and enables focus on the functional rather than necessitating identification of every possible reliability failure in every sub-component. Abstracting to function enables attention to hardware, software and liveware within consideration of the points of control. Nothing needs to be broken for negative outcomes to occur. There is a methodology and it is traceable, permitting third party review of the logic in deciding what is in and what is out. But more importantly, there is no need to identify every reliability failure mode of every sub-component, so the concept of a robust analysis isn't in question. By adopting a functional abstraction and analyzing points of control, the idea of missing something becomes a little odd until recognising that the question assumes decomposition to components. Distilling this into one paragraph doesn't do it justice. We owe the systems safety community a debt of gratitude for driving system reliability to ten to the minus nine rates. But a traditional systems safety approach will never address failures of emergent functions, as these will never be part of a sub-component decomposition. Without reliable components, systems don't function long enough to become complex—to be dynamic and feature emergence. They are just unserviceable. A tool like STPA is different and comes after you've achieved reliability with FHA, FMECA, ETA, SMS, TLA, ETC.

Kay and King (2020) do a bang-up job of explaining the difference between probability and uncertainty. Well worth the read – it taught me to be comfortable with uncertainty, without any need to put a number on it. Probability is a ratio and knowledge of a probability requires knowledge both the number of outcomes and the universe of possible outcomes. The number is meaningless when you don't know the denominator in a probability.

What I'm doing is abstracting and codifying what flight test already does in our best practice, articulating the reason behind why we should listen to our cultural lore, and sharing that with the wider world.

Cheers,
Ben
benjamin.luther@adelaide.edu.au

Letter to the Editor: "Pancho Weighs in"

Sarah "Pancho" Summers

Editor's note: I reached out to Sarah for comment on my STPA Deep Dive editorial prior to publishing it, and unfortunately I used the wrong email address and didn't provide enough time for a response. But when she finally did see my inquiry, Sarah was kind enough to respond and share her slides.

Mark,
First off, sorry for the delay in responding! I hope this can still be of use. I am in school right now, so I'm not checking this email on a regular basis.

I agree that STPA can't be self taught, and that definitely is an area where the STPA community can continue to evolve. Prof Leveson conducts an annual free STPA workshop that can be attended both in person and virtually. Anyone who is interested in learning more definitely should check her website and consider attending next year. One of the reasons STPA takes a while right now is because it's not used in the design phase. Therefore testers have to take a lot of time to develop the model and then conduct the analysis. If design engineers incorporated STPA into their processes and it was handed to testers all testers would need to do is add the test specific aspects to the model and run the resulting delta analysis. While STPA takes longer on the test side compared to traditional methods, it is a lot faster on the design side and because it focuses on system behavior, it can be readily adapted to a flight test program. Testers seem to get the value of STPA, but we haven't had as much success pushing the rope on the design side. I think if we could get designers on board this would be a game changer.

Regarding probabilities: it helps to think about what the purpose of using probabilities in a safety analysis is. I see probability in a safety analysis as twofold: two determine what the likelihood of bad outcomes are so the test team can focus on the most likely outcomes & to communicate the risk to approval authorities. Are there other ways to both determine where to focus and to communicate? I think STPA can provide avenues for both. For focusing efforts, you can use both traceability to a hazard that is more severe, and you can use how many scenarios a particular mitigation may solve (I call this low hanging fruit). For communication, I have been in many test programs where we make our best guess at probability (severity is much easier to determine), but we find we realize hazards more than expected due to our initial lack of system behavior knowledge. However, because our approval process is based on severity and probability, we may have done inadequate safety planning & approval. Also, because we use some number, 10^-9 as an example, it may lead people to believe there was more analytical rigor to the probability analysis. Lastly, the functional diagrams are a great way to communicate function and potential risks to a safety board or approval authority. It ensure everyone has the same mental model of the system under test.

To be clear: I am not anti-probability. However, there are a lot of system behaviors that can't be modeled using statistics that we often try to shoehorn into a probabilistic methodology in a way that could lead us down a bad path. As I believe I mentioned during the talk, I think overall we are very good at safety, particularly from a mechanical reliability perspective. We have an opportunity to continue to use those methods, but also look at system behavior

in a different way to get after the challenges we often face now with complex software, system interactions, and human/machine interactions. Please shout if you'd like to talk about any points I've made.

Sarah
sesumme@gmail.com

Turbo Talk

Art "Turbo" Tomassetti

Hello everyone. It is back to school time—well not for most of us probably, but we remember what that was like. Seeing old friends, making new ones, sports, activities...oh and of course learning stuff. Think back to what you thought about subjects like Chemistry, Algebra, Latin, etc., before you took the first class. Then think about your thoughts on those subjects after you had completed the class. Better? Worse? Meh? No matter what it was, it was more informed. Which is my segway into this edition's discussions on STPA. Personally, I have had minimal exposure. Dr. Nancy Levenson was an early guest on our podcast, and we have offered sessions on STPA at several Flight Test Safety Workshops. I have never had a chance to actually practice it in the real world, but I know that my thoughts on it are different now than when I first heard the acronym. So, if you are new or mostly new to STPA give the articles in this edition a read and then go dig a little on your own. Check out episodes 7 and 8 of our podcast or take a look at what we have on our website and develop an informed perspective.

Speaking of our podcast this month, I talked to another Marine friend who shares an edge of your seat kind of story from his F/A-18 days. One of the things we touch on in the discussion is the use of simulators for emergency procedures training. In my military days I pretty much had regular access to a simulator, at least when not deployed. But I rarely went to the simulator any more than I was required to. I know—shocking that on any rainy day I didn't say, "Hey I want to go run emergencies in the sim for an hour or so". But during our interview I learned my guest, now a business jet pilot, gets to the simulator only once or twice a year. Awesome right? Who wants to go to that thing anyway. But flight simulators are a great training tool. You can practice a lot of things in them, and we all know what practice does for your skill and proficiency. And consider that maybe the best way to recognize and deal with danger is to go practice danger. IN the safe environment of the flight simulator of course. Anyway, be sure to check out this month's podcast and I welcome your thoughts and experiences on the use of flight simulators.

Until next time: Be Safe, Be Smart and Be Ready. **Turbo**

Latest Podcast

Art "Turbo" Tomassetti

EP 45 – Are You Ready For the Bad Day? Part One - The Water's Getting Closer — You can subscribe to the Flight Test Safety Channel podcast in iTunes, Spotify, Podbean, Google Play, and Amazon Music's FTSCChannel. You can also share the link below with colleagues and friends who may not know about Turbo's monthly recording and navigate directly to the podcast: https://flighttestsafety.org/ftsc-news/flight-test-safety-podcast-channel.

Contact the *Flight Test* Safety *Fact*

Mark Jones Jr, Editor
mark@flighttestfact.com

Art "Turbo" Tomassetti, Chairman	chairman@flighttestsafety.org
Susan Bennett, FTSC Administrator	susan@setp.org
Society of Flight Test Engineers	edir@sfte.org
Society of Experimental Test Pilots	setp@setp.org
AIAA Flight Test Group	derek.spear@gmail.com

Connect with us by joining the LinkedIn Group: "Flight Test Safety Committee."
Website: flighttestsafety.org