Safety Insights

Focusing on the human contributions to risk

Victor Riley

1. Introduction

I spent the last several years of my career in aviation human factors trying to answer the following questions: how can we better anticipate, analyze, and assess human-related risks when there are too little data for traditional statistical analysis? When we learn of a serious incident involving a surprising human behavior, how do we determine whether it presents a potential risk to safety?

By human-related risks, I mean unexpected human behaviors that either create hazards or fail to mitigate hazards that have otherwise emerged. For example, pilots are supposed to be trained to mitigate a wide range of hazards, including problems with the airplane or environmental threats. But sometimes, they may act in unexpected ways due to a variety of factors:

- an instinctive reaction may hijack the behavior away from the expected trained response;
- a key training or skill gap may prevent the pilot from responding as intended by the designers;
- an erroneous belief or assumption could cause a pilot to misinterpret the situation.

Any organization that designs, builds, or operates complex systems is susceptible to expecting the people on the front lines of using or operating them to do so as the designers intended. But incidents and accidents in all domains occur when people act in unexpected ways. When we see this occur in an incident, how do we evaluate a newly recognized risk to decide if it poses a potential safety issue? There are lots of methods and tools available to help with this, but sometimes they can mislead rather than inform.

The European Union Aviation Safety Agency (EASA) recognizes the importance of assumptions and evidence that may indicate deviations from them in a safety bulletin to airlines, which says in part: "To enable the in-depth analysis of in-service events involving human interventions, the assumptions, which have been made by the DAH (Design Assurance Holder, aka the manufacturer) when demonstrating compliance with the certification basis about the expected flight crew behaviour, need to be known in order to identify any deviations from these assumptions in the context of operation." The bulletin goes on to recommend that airlines report any unexpected pilot behaviors that may lead to hazards to the manufacturer because they can't fix what they don't know about.

This is a critical part of improving aviation safety because, in the vast majority of recent accidents, unexpected pilot behaviors were either causal or contributing.

So how do you know when a surprising behavior in an incident could potentially lead to an accident someday? How do you associate it with a safety threshold? In order to understand this better, I read a number of books on risk, probability, and prediction. Why prediction? Because any decision about whether a system or condition should be considered safe or potentially unsafe is a prediction. When a product is certified or approved, everything from airplanes to drugs to medical equipment to cars to power plants and beyond, that approval is itself a prediction that the product should be considered safe. So the literature on how people make predictions and what kinds of biases can affect their accuracy is relevant to assessing safety risks.

The books I read include:

- The Incerto series (Fooled by Randomness, The Black Swan, Antifragile, and Skin in the Game by Nassim Nicholas Taleb)
- Bernoulli’s Fallacy by Aubrey Clayton
- Expert Political Judgment and Super Forecasting by Philip Tetlock

- The Signal and the Noise and On the Edge by Nate Silver
- How to Measure Anything by Douglas Hubbard
- The Yellow Pad by Robert Rubin

Anyone who has read these books could come to the same conclusions I have. But since these books aren't intended to specifically relate to safety, and I haven't seen anyone else make the (I think valuable) connections to it, I do so here.

These essays relate to domains, particularly aviation, where serious incidents are rare, catastrophic events are exceedingly rare, and the standard for safety is exceedingly high. In aviation, that standard is typically 1E-9 (or one in a billion chance of a catastrophic outcome), and the known incidents that may be precursors to such an event are usually in the single digits if they exist at all. With such little statistical power and such a high standard of proof, other methods are needed. And since the human contribution to a serious event is often complex and unpredictable, this is where such methods are needed most.

I will not be referring to any incidents or information from my prior employers, so this will be a general philosophical treatment with no specific examples. However, I expect that an informed reader will think of their own examples.

The contents of this website can be downloaded as a pdf document.

2. Some models are harmful

The British statistician George Box famously said, "All models are wrong but some are useful." This overlooks the possibility that some models can actually be harmful.

In The Black Swan, Taleb refers to a Lehman Brothers employee who claimed, in 2007, that the developing housing crisis was a "one in ten thousand year event." Now, which is more likely? That it really was a one-

in-ten-thousand-year event? Or that the models that said it was were wrong?

Because it was those very same models that enabled the crisis in the first place. Those models, used to price derivative financial products from bundles of subprime loans, predicted that the loans were independently exposed to failure and the expected failure distribution would follow a normal bell curve. But, in fact, the underlying loans were highly coupled because they were almost all exposed to the same set of risks:

- A large set of borrowers who had been encouraged to take on the most debt they could qualify for;
- Mortgage issuers with lax borrower qualifications standards;
- Regulators who rated the risks of bundled assets to be prime when the underlying assets were subprime;
- A rising interest rate environment that caused many loans to fail all at once.

Finance companies had such faith in the models that overlooked these dependencies that some of them leveraged their own assets by thirty or forty times in order to amplify the profits they were getting. And when the loan failures started cascading, this leverage amplified their losses to the extent that several large firms, including Lehman Brothers, failed as well.

Why start with this topic? Because a model is a set of beliefs, assumptions, and expectations about the world: in other words, a theory about it. And every organization that designs, builds, approves, operates, or administers products necessarily has expectations for and assumptions about the end users of those products, and when that organization tries to assess the risk revealed by a serious incident, they'll likely do so using their existing expectations and assumptions.

Expressing these expectations in formal models can give organizations a false sense of security that they fully understand the issues and have contained the risks. This can cause them to focus on managing the risks

they recognize and stop considering whether they've really even recognized them all. And this is why so many accidents are surprising: because they reveal risks that previously weren't recognized.

For example, you might think that a casino would be the most competent organization to manage risk, since that's essentially their business model. So what are the existential risks a casino might face? A large lucky bet by a high roller? A highly improbable run of outcomes against the house?

Taleb points to some of the truly existential risks casinos have experienced but hadn't anticipated:

- The tiger attack on performer Roy Horn which was estimated to cost the hosting casino in the range of a hundred million dollars;
- A casino owner's daughter who was kidnapped and held for ransom;
- A disgruntled contractor who tried to blow up the casino he had worked at;
- An administrator who was supposed to file tax forms with the IRS but instead filed them in his desk, exposing the casino to the loss of its gambling license and large fines.

I suspect that many accidents in transportation, process control, medicine, and other domains that were ultimately attributed to human error were similar surprises to their industries when they happened, because the expectations in those industries didn't account for those "errors". And those expectations are based on what they think their risks are and how they model them. But once those surprises happen, the next important thing is to decide whether to dismiss them as anomalies and persist in one's prior beliefs, or to learn from them. In other words, which do we believe more: the model, or the reality?

3. A few thoughts about human error

It's widely recognized today that most airplane accidents and most hospital deaths are due to human error. And in a purely technical sense, where the

word “error” refers to deviation from the desired path, this is accurate. But the common use of the word “error” presumes that the person or people involved are fully trained and competent, and that they merely made a mistake. In many cases, these attributed “errors” may not be errors at all, in that sense.

For example, if I try to play a Paganini violin concerto and screw it up because I’ve never played the violin before, that’s not an error. Likewise, if someone is involved in an incident or accident because they either didn’t mitigate a hazard or perhaps even caused it, it might be because:

- they lacked the training or skill needed;
- they had an erroneous belief about the situation or an invalid assumption that defeated whatever layers of protection assumed a proper understanding of it;
- they couldn’t resolve or were misled by conflicting cues;
- their training and/or experience predisposed them to react in a way that wasn’t appropriate for that particular situation.

Each of these has caused serious incidents and accidents, but none of them can really be considered an “error” in the common sense. Instead, I propose using the term “adverse action”, which simply describes the fact that whatever the people involved did either didn’t help the situation or made it worse. It’s also less judgmental than saying that they screwed up.

4. Human operators are really people

In human factors and human engineering, the term “human operator” is commonly used. As a dry technical term, it suggests a human component to a system that functions in a predictable and reliable way, like other hardware and software components. I wonder if the broad use of this term predisposes engineers to assume that the “operators”, once trained and qualified, can be treated analytically like any other system component.

So it's worth remembering that human operators are actually people, and therefore subject to all the faults and frailties of any person. This includes surprise, stress, workload, divided or channeled attention, distraction, cognitive biases, jumping to conclusions, reacting instinctively, and many others. And human performance is ultimately the product of human behavior, which in turn is driven or shaped by an almost infinite set of influences. These include:

- personality characteristics
- culture
- experience
- all of the specific circumstances (environmental, system state, personal, etc.) present in any given situation.

In domains like aviation, the shear number of variables that can affect what people do in a specific situation is so large that it's impractical to try to predict outcomes probabilistically. Since no amount of training turns people into machines, the human is always the most complex and least deterministic part of any non-autonomous system.

5. Probabilities and possibilities

This is why it's usually impractical to predict human behaviors in complex environments such as aviation probabilistically. Not only are there an unbounded number of variables that might influence behaviors, but which of those variables will weigh more heavily than others is also unknown. While you can take a piece of equipment into a lab and subject it to stressors until it fails, you usually can't do that with people. And while the range of variables that can affect equipment is usually small (temperature, vibration, cycles or hours of use), the range of variables that can affect human behaviors can't be tested to any acceptable statistical power. Trying to put a number on human behavior probabilities invites the overconfidence in a model discussed earlier.

Furthermore, trying to predict human behaviors probabilistically can tempt one into favoring expectations over evidence (or models over reality). If someone does something unexpected in an incident or accident, it's tempting to dismiss it as an anomaly and continue to believe in the prior expectations. In these cases, one could examine the behavior and try to explain it in order to determine if it actually presents a systemic risk, or one could dismiss it until enough of them happen to persuade a skeptic that it's real. Since a statistical analysis requires more than one data point, and every accident (in aviation, at least) is unique, it's illogical to try to extrapolate a trend or make a probabilistic prediction from it.

Taleb makes the distinction between favoring theory over evidence in his distinction between economists (favoring theory) and traders (favoring evidence). With no real skin in the game, it's easy for economists to carry theoretical expectations beyond the point where they're no longer valid, while traders are penalized if they ignore immediate evidence. (This may be why it took economics so long to evolve from assuming everyone was a rational, optimizing economic actor to recognizing the tenets of behavioral economics.)

As an example of this, he asks us to imagine a fair coin being tossed 99 times with the result being "heads" on every toss. What's the probability of getting "tails" on the 100th toss? A frequentist would say it's 50%, because the coin is fair and every toss is independent, and anyone who thinks otherwise is subject to the "gambler's fallacy". But the evidence is showing that the coin really isn't fair, and the actual probability of "tails" on the next toss is zero. (A Bayesian calculation would confirm this.) The statement that the coin was fair sets up an assumption that the evidence is contradicting, and revising one's conclusion requires recognizing and questioning one's assumptions.

To me, this has deep implications for how we should treat evidence from incidents and accidents. Even if there's only one unexpected event, it probably has something to tell us. It shows us a new possibility, or demonstrates that something we may have thought highly improbable is actually plausible enough to be taken seriously.

6. Probability and plausibility

The debate between the two primary schools of probability, frequentist and Bayesian, has of course been going on for over a couple of centuries. Both have their shortcomings when applied to safety.

For example, the frequentist approach applies to large numbers of events and a frequentist probability is a prediction of how often something is expected to occur over a number of opportunities for it to occur. Using frequentist statistics and associated probabilities requires a subjective decision about significance levels (that is, when do you take a result seriously), and they require enough relevant events to indicate a trend. Obviously, when there's a surprising accident or serious incident, it's not in the interest of safety to wait for a significant trend to emerge.

In contrast, a Bayesian approach is better suited to rare events. While a Bayesian analysis would do a better job of incorporating and learning from smaller numbers of new events, updating a prior probability requires that there be a prior probability in the first place, which is also necessarily subjective when dealing with rare events. In other words, one has to have a theory and be willing to update it based on new evidence.

According to Aubrey Clayton (Bernoulli's Fallacy), recent attempts to reconcile the two schools have focused on the notion of "plausibility". He summarizes this work by saying that "probability is best understood as the plausibility of some event given some assumed information". Note that plausibility can apply equally well to expectations about the frequency of occurrence and about the likelihood of single outcomes.

I think this is a good construct to use in safety analyses of rare events, and that the safety-related standards and literature already tacitly use it. For example, the global airworthiness regulations refer to "any foreseeable operating condition" and "reasonably expected", which are just other ways of saying plausible.

And I think it's particularly useful when dealing with human-related risks because it helps bound the analysis space. Human behavior is typically not random. While it is highly variable and impractical to predict, it's at least systematically related to the combination of human characteristics and

circumstances. The whole field of human factors has been dedicated to understanding these relationships, so human factors can help us separate plausible scenarios and outcomes from those that are implausible.

7. Analysis methods

So now that we have an idea of what criterion to use, how do we get there? What kind of analysis methods make best use of it?

There's been a lot of interest lately in STAMP/STPA (Systems Theoretic Accident Model and Processes/Systems-Theoretic Process Analysis), developed by Nancy Leveson at MIT. And also the Bow Tie analysis method, which is intended to help identify safety threats, barriers that should prevent those threats from becoming hazards, and if a hazard occurs, barriers that should prevent the hazard from resulting in an accident.

The original Bow Tie method was intended to be the combination of a fault tree on one side and an event tree on the other. The fault tree described the potential failure modes of a system, and the event tree was used to map out how a hazard resulting from those failures could evolve to outcomes. Over time, it's been modified into a purely qualitative barriers analysis on both sides. This is very well aligned with James Reason's Swiss Cheese model of accident causation (more about that later).

Leveson has published an insightful critique of drawbacks of the Bow Tie method, one of which is that the method presumes that the analyst already knows all the potential threats and hazards. (Fault trees also have this issue). However, while STPA provides a much better analysis of where control structures exist and feedback could be needed, I think it also starts from step 2 rather than step 1.

So what is step 1? To me, it's a more comprehensive exploration of how an initiating condition might develop into all plausible outcomes, which is best done with an event tree. The purpose of this is to map out what did happen, what should have happened, and what else could have happened instead. This event tree should be informed by human factors analysis so plausible alternative actions can be considered at every step. It would start with the initiating condition and map out all the plausible pathways, both safe and

unsafe, based on known human tendencies and performance shaping factors. In my experience, this exercise has revealed previously unrecognized hazards and associated pathways to them. It can also help identify key points where the condition may be fragile to particularly dangerous but plausible decisions or actions. This can then point to where additional barriers (feedback, protections, etc.) may be needed.

Referring back to “Some models may be harmful”, I suggest that this type of model is intended to be exploratory rather than explanatory, to raise questions rather than to nail down decisions. Filling out as complete an event tree of plausible pathways based on human factors principles should, in my view, precede STPA, Bow Tie, or any other method that’s intended to help inform a safety decision. And to me, that's the proper role of a model: to make sure you've thought of everything. The problem with the finance industry leading up to the 2008 housing crisis was that they were letting the models think and decide for them.

8. Swiss Cheese?

Just as “all models are wrong but some are useful” has perhaps blinded us to the possibility that some models may be harmful, I think the broad adoption of James Reason’s Swiss Cheese model has predisposed us to think that all accidents result from the failures of multiple independent barriers.

However, just as the financial models in the 2008 housing crisis failed to account for common-cause loan failures, I think there are cases where the Swiss Cheese model doesn’t hold. I think these cases are characterized more by key points of fragility than by combinations of failed barriers.

For example, an erroneous assumption can defeat multiple barriers at once. So can the absence of an expected skill or a key training gap. An initial misinterpretation of a single prominent cue may cause a person to overlook or dismiss other cues that tell a different story. Someone who is poorly equipped to mitigate a minor hazard is probably even less well equipped to mitigate the major hazard it evolves into. In these cases, the barriers are coupled rather than independent, so their combined effectiveness should not be treated as the product of independent probabilities.

So one of the goals of an event tree is to reveal those decision or action points that are potentially susceptible to common human failings, like jumping to the wrong conclusion, reacting instinctively rather than deliberately, persisting with an unproductive action, inattention to the key cues, expectation bias, and so forth. Any competent human factors expert can undoubtedly come up with a list of potential influences that may drive behaviors toward an unsafe path. Finding those key points of fragility can then help us understand how an adverse response may lead to a catastrophic outcome, and suggest ways we can prevent it.

It's true that many, perhaps most, accidents are well described by the Swiss Cheese model because multiple barrier failures stack up to allow them to happen. But I think there are some accidents that are better described by single points of fragility that set the event onto a hazardous path in which none of the intended barriers can be effective.

9. Probabilities and dependencies

When thinking about probabilities, even in a qualitative sense without numbers, it's critical to distinguish between absolute probabilities and conditional probabilities. Even if we think in terms of whether something is plausible rather than probable, as suggested by Aubrey Clayton, plausibility in safety usually depends on context and must be considered as conditional rather than absolute. For example, an instinctive reaction may be the right response to a condition in most cases, but is there a case where it's exactly the wrong response because of slightly different conditions?

It's also important to remember that redundancy works differently with people than with hardware or software. One may achieve full redundancy with independent dissimilar systems, but adding people or redundant steps to a task or procedure doesn't achieve the same effect because those additions are not fully independent.

One reason is due to the diffusion of responsibility, a term from psychology that describes the fact that when multiple people are responsible for a task, it's tempting for each one to pay less than full attention to it because they assume someone else will catch any errors. This also applies to humans and automation: there have been many accidents in which people assumed automation would handle a situation when it couldn't.

Trying to strengthen a task by adding steps that a single person has to perform, like checking or double-checking that something was done right, is susceptible to the same problem: the assumption that "I'll check it again later so I don't need to pay too much attention to it now," or, "I've already checked it so I don't need to pay as much attention this time." This is where expectation bias and confirmation bias come in. Adding more opportunities to catch an error quickly reaches a point of diminishing returns because expectations cause those opportunities to be coupled rather than independent.

For all these reasons, when adding people, automation, or steps as barriers to catastrophic outcomes, it should be recognized that any dependencies between them will prevent them from being fully effective.

10. Predictions

When we analyze a condition to decide whether it presents a safety risk or not, we're essentially making a prediction about its potential future outcomes. This is also true when a regulator approves a product under the belief that it has been satisfactorily proven as safe. The people making these decisions are subject to many of the same influences that can affect how people operate systems: expectation bias, confirmation bias, channeled attention, erroneous assumptions or beliefs, etc. So I think it's useful to think about how people in the organizations that design, build, approve, and operate complex systems decide whether they should be considered safe or not.

Two of the biggest influences are the personalities and internal politics of those organizations, and here, the work of Philip Tetlock is relevant. Tetlock is a psychologist who, several decades ago, got interested in the question of how good the predictions made by various pundits on TV actually were. After all, the role of an expert pundit is typically to predict how the news of the day will drive the news of tomorrow. To study this, Tetlock ran a multi-decades-long experiment with over a thousand people to measure how well they made predictions, both within and outside their professed areas of expertise, and their personality traits.

To summarize his findings at a very high level, he distinguishes between two primary personality types: foxes and hedgehogs (following an ancient

Greek saying that, “The fox knows many things but the hedgehog knows one big thing.”) Foxes tend to seek out information from more diverse sources and are more self-critical and therefore less self confident in their predictions. Hedgehogs consider themselves experts and are less self-critical and more self confident in their predictions. Where foxes may waffle, hedgehogs are decisive. The “one big thing” the hedgehog knows is a global theory through which they interpret events and make predictions. Examples in the world of punditry include neoconservatism and trickle-down economics.

The elevator speech version of Tetlock’s most interesting finding is this: in areas of their expertise, the more a fox knew about their subject, the better their predictions. But the more a hedgehog knew about their subject, the **worse** their predictions. Since hedgehogs tend to be more confident, one implication is that the more confident you are about your predictions, the more likely they are to be wrong.

The reason for this is that where foxes favor evidence, hedgehogs favor theory. Their commitment to a strong, global theory of the case drives their predictions to the exclusion of countervailing evidence. The more they rely on their worldview to drive their predictions, and the more they interpret new evidence through the lens of that worldview, the worse their predictions are.

So who gets on TV? The fox with their caution and caveats, or the firm, decisive, compelling, and self-confident hedgehog? Of course, it’s the latter, and this is why pundit predictions are almost universally unreliable. It turns out that most pundit predictions are wrong.

Tetlock doesn’t extend this thought to organizational dynamics, but I think it applies. Depending on organizational culture, the self-confident hedgehogs may advance into leadership positions while the more cautious foxes get left in the trenches. I think this may partly account for the common disconnect between the lower and upper layers of large organizations: why the lower layers often resent the upper layers, and the upper layers often discount the lower layers.

I suggest that it’s very useful for anyone involved in safety in a large organization to be aware of this dynamic and be able to identify when a

strong personality may be driving a decision based more on their own beliefs and expectations than on all the available evidence. I think a hedgehog is more likely to dismiss a surprising, unexpected action in an incident or accident as an anomaly because their global theory doesn't allow them to accept that such an action should have been possible, let alone that it might happen again.

11. Safety policies and procedures - the weak link

Why aren't policies and procedures intended to maintain safety more effective? I suspect it's because they compete with other cultural values.

For example, a worker who shortcuts a safety procedure at a process control plant because it takes too much time is prioritizing an immediate, tangible, and rewarded cultural value (efficiency) over an abstract policy intended to prevent a rare and unexpected outcome. In this case, the tangible value of saving time and effort outweighs the real cost of following a less efficient procedure when the real outcome is almost certain to be the same in either case. Then the exceptional tragic outcomes raise questions about why the procedure wasn't followed.

I think that safety policies and procedures will continue to be weak barriers unless they themselves are tied to a recognized and widely shared cultural value. I suggest that value be ethics. After all, safety is ultimately an ethical issue. The people involved have a fiduciary responsibility for safety, and if the organizational culture can associate following safe practices to the ethic of protecting oneself and others, maybe that would help balance out the competing value of efficiency.

This extends to how an organization deals with surprises that challenge their expectations and beliefs. Does that organization value humility and learning? Is it ready to acknowledge failures and improve? Or is it so committed to its traditional expectations of how the world should be that it rejects contrary evidence about how it actually is? If a learning culture is a safety culture, it requires recognizing and questioning one's own assumptions and being ready to learn when they may not be valid. And if safety is an ethical issue, the humility and curiosity needed to question assumptions and learn should be as well.

12. Putting it all together

So, in the end, how do you evaluate whether a surprising human behavior that is revealed in a new incident represents a potentially serious threat to safety? I recommend the following steps:

Build a full event tree, mapping out all the plausible pathways from the initiating event or condition to outcomes. At each step, consider what the person or people involved may do, including what we would hope and expect and what they might do instead based on known human factors considerations. Consider the potential for instinctive reactions, negative transfer from previous experience, common errors, relevant cognitive biases, and so forth. This will help fill out the picture of how the initiating condition or event could potentially evolve into a catastrophic outcome.

Along each pathway that leads to a potentially catastrophic outcome, identify the barriers that are expected to prevent such an outcome. Consider feedback, procedures, alerting systems, and other protections. Then rate the expected effectiveness of each barrier. In general, barriers that rely completely on human behaviors should be considered weak. Also consider whether the identified barriers are completely independent of each other and can therefore be considered fully redundant, or whether there are dependencies between them that would make them less than fully redundant. I recommend this step instead of a bow tie analysis because putting the barriers right on the event pathways puts them in the appropriate context and allows them to be depicted sequentially if one backs up another. This helps clarify their conditional relationships, where a bow tie analysis may lead one to expect them to be independent.

Based on a qualitative assessment of the likelihood (not probability) of the human actions that lead to each hazardous pathway and the combined strengths of the barriers between those actions and catastrophic outcomes, decide whether that particular pathway or sequence of events tells a believable story about how a catastrophic outcome could occur. The objective here is to determine whether avoidance of a catastrophic outcome can be reasonably assured. If there are key points of vulnerability or fragility (decision or action points where known human factors influences make an adverse action likely), consider whether additional barriers, such

as more salient feedback or new protections, could strengthen the barriers that would either prevent the action or allow graceful recovery from it.

If a formal decision about whether the condition should be considered a safety hazard is needed, your organization may need to formalize an acceptance criterion to make that decision. Again, I recommend that it be qualitative rather than quantitative, but these analysis steps should provide a framework for defining a criterion that's appropriate for your industry and role in it.

Of course, this is likely just the first analytic step in a longer process that may involve human-in-the-loop testing and possibly the development and testing of corrective actions.

This process is only as effective as the people doing it. Since human behaviors are the keys to the various potential pathways, the event tree should be built by experts in both the domain and in human factors. The team should be mindful of their own prior expectations and potential biases, and recognize and be willing to question their prior assumptions. Ideally, the team should be able to avoid the temptation to say, "No one would ever do this," without strong rationale to justify it. A search of prior incidents that may be precursors to the surprising behaviors can be useful here to provide either confirming or contradicting evidence.

13. A proposal to improve aviation safety

If you follow aviation, you've probably been surprised by the pilot actions that led to some high-profile accidents. You might have never expected a pilot to raise the nose in response to stickshaker as the Colgan Air 3407 pilot did in 2009. You might have never expected an experienced transport pilot to crash a 777 during a visual approach into a large airport on a clear day as the pilot of Asiana flight 214 did in 2013. When you're surprised by such an event, it's because that event violates your assumptions.

Manufacturers have to make assumptions about line pilot training, skills, and the qualifications that are supposed to defend those skills when making design decisions, writing their documentation, and performing their safety analyses. But today, there's no end-to-end continuity from those assumptions to end-user line pilot training. Training is usually provided by

the carrier (either in-house or outsourced) and approved by the carrier's local regulator, and the manufacturer has no way of knowing whether that training meets all of the design and safety analysis assumptions.

Between 2013 and 2022, I count 14 hull losses just in the Boeing fleet in which unexpected pilot actions were either causal or contributing. In only four cases was anything wrong with the airplane, and most of those fourteen accidents involved pilots with more than 10,000 hours. During that time, there were no hull losses due to systems or structures failures that the pilots couldn't mitigate. And Boeing is not alone - there have been similar accidents in the fleets of other manufacturers. I believe that the discontinuity from manufacturer assumptions to line pilot training is the biggest threat to aviation safety because it underlies the vast majority of accidents.

The US Aircraft Certification, Safety, and Accountability Act (ACSAA) requires the FAA to validate the pilot assumptions used in design and certification, and of course the FAA flows this requirement to the manufacturers. There are only two ways to do this: either gather enough information from the worldwide fleet to detect emerging training and skill gaps before they cause accidents, or make sure those gaps don't emerge in the first place. The second is far more practical and effective.

To do this, I propose that manufacturers provide operators (airlines) and their local regulators lists of the critical training assumptions or requirements that support the design safety analyses. Regulators around the world must then commit to enforce those requirements. Thus, when an operator requests approval for a training program that omits training that the manufacturer assumes or deems to be critical, the regulator can consult the list and deny that request. In other words, the manufacturers should tell the industry what their expectations are, and the industry should make sure those expectations are met.

Final thoughts

In this series, I've tried to bring together ideas from finance, psychology, risk management, probability, analysis methods, prediction, and human factors to gain insights into better anticipating, analyzing, and assessing human-related risks. The common theme throughout is the tension

between expectations and evidence. One of the things that makes an assumption an assumption is that it's taken for granted - difficult to recognize let alone to question. But this is one of the most important things anyone working in safety can do.

The most difficult assumptions to recognize may be those related to human factors. It's tempting for anyone to believe that, because they're human, they're able to predict how other people will react to situations. But people systematically overestimate the extent to which other people think like they do, and it takes a certain level of enlightened awareness of this tendency for people to resist it, to recognize their assumptions and acknowledge that they may be wrong. Every surprising incident should point to such an opportunity; it surprises because it violates an assumption or belief.

I hope these insights will help you and your organization manage these risks better.