LESSONS LEARNED AND MURPHY’S COROLLARY
By Michael Meier

Abstract:
Safety in flight test requires that when the potential consequences of a negative outcome are catastrophic, the go decision must be made only when a very high degree of confidence in a positive outcome exists. When this protocol is followed, it is a natural result that a very high percentage of outcomes will be positive, and it is further likely that we will see a long string of consecutive positive outcomes. When this happens a feedback loop is created in which the go decision is validated by the positive outcome as having been a good decision, and is therefore reinforced and made more likely to be made again in the future under similar circumstances. Beyond that, often the lesson learned from a series of consecutive positive outcomes is that our evaluative criteria for making the go decision were unnecessarily conservative; not only were we correct to make the go decision, but we could have made it under even somewhat less favorable circumstances. This can lead to a gradual lowering of our standards.

However, consecutive successful outcomes can mask go decisions that may be fundamentally unsound, if the success is the result, either partly or entirely, of good fortune. Below a certain threshold of probability for success, repeated go decisions will lead eventually to an unacceptable probability of failure. At 99% probability of success, eighteen events results in a cumulative probability of failure for at least one event of 17%, or one chance in six.

The common idiomatic statement of Murphy’s Law is: “Anything that can go wrong will go wrong.” This is likely not what Murphy said, nor is it the best statement of the lesson learned from Murphy’s experience, and in fact, no one who thinks about it can actually believe it. Almost anything could go wrong, and if it were really true that anything that could go wrong would go wrong, then most everything would go wrong, and most things don’t. But if Murphy’s Law were true, then a corollary of this “law” would also be true: “Anything that did not go wrong could not have gone wrong.” And while we don’t really believe Murphy’s Law, we are surprisingly susceptible to being seduced by an unconscious belief in Murphy’s corollary – that is, we take a positive outcome as evidence that a positive outcome was inevitable.

Combining with our failure to appreciate the nature of probability and our susceptibility to Murphy’s corollary is the inevitable and powerful pressure we often experience to make the go decision under less than favorable circumstances. This pressure is a natural and unavoidable consequence of project deadlines, budgets, and performance expectations. It induces us to make the go decision when we otherwise might not, and to validate the go decision afterwards when the outcome is successful, when that validation may be inappropriate. A means for combating the effects of these three factors is to conduct a rigorous and quantitative pre and post-test analysis not only of failed outcomes, but also of successful ones, with the goal of identifying apparently successful outcomes that may have followed bad decisions, and that were successful because of good fortune.
Background:
The ideas in this paper came out of thinking about a serious hang gliding accident in 1995 that occurred on landing in circumstances under which hundreds of previous successful landings had been made. The ideas were first expressed in an article in Hang Gliding Magazine published in 1998. The article can be read here: https://www.willswing.com/why-cant-we-get-a-handle-on-this-safety-thing/. They were adapted for presentation to SETP and presented at the 2015 West Coast Symposium. Because the material here was developed for verbal presentation, this paper essentially follows the form of that presentation. Some additional material in the way of amplification and background that could not be included in the presentation due to time constraints is provided. The added material is in italics and indented.

LESSONS LEARNED AND MURPHY’S COROLLARY

Sharing lessons learned, with the goal of enhancing the effectiveness and safety of experimental flight test. That’s why we’re here. So I want to pose a question: Might we be missing something in this process? I’ll suggest an answer a little later on, but right now I’d like to invite you to participate in a thought experiment.

Imagine that you find yourself in need of a significant amount of money, with no obvious way to get it, and with serious negative consequences if you don’t. I come along and offer you ten million dollars, if you will play a single game of Russian roulette. One bullet, six chambers, one pull of the trigger. If the gun fires, you’re dead. If not, you get the ten million, and your problems are over. OK, I realize your problems would be over in the event of either outcome. Let’s disregard that for now. So, the question is, "Would you play?"

I work for a company that designs and manufactures hang gliders.
For 39 years, I’ve been engaged in the experimental, developmental and production test flying of the gliders we make. Production test flying is the final quality control check for a manufacturing process that is part science and engineering and part art and craftsmanship, and we’ve found it to be an essential tool in maintaining manufacturing consistency.

Profit margins in hang gliding are pretty thin, and like most businesses, we continually look for ideas to reduce cost and increase efficiency. About 35 years ago, we hit upon one such idea for production test flying.

Normal procedure involves launching from the top of the mountain, conducting the flight test, and landing in the landing area in the valley below, then driving back up for the next round – a process which might be repeated as many as eight times in a day.
To save time, we thought we could thermal up from launch, perform the flight test, and then land back on top near where we took off from, saving ourselves the drive back up from the landing area and eliminating the need for one crew member - the driver who would otherwise be needed to drive the truck down the mountain to pick us up after each round.
Not every hang gliding launch site lends itself to top landing, but one of ours did.

Still, the landing demands precision and the right approach.
Try to approach into the wind and there are two likely outcomes – either you’ll be too high or too fast and overshoot, or you’ll be too low or too slow, and descend into the lee side rotor behind and below the hill and lose control in the turbulence. The window of “just right” between the two is extremely small.

So we developed an alternate approach – coming in half crosswind, more from the side of the hill than from the back, approaching with good speed and landing on the upslope, slightly below the top.
Continuous turn through base to final

Flaring for landing on the upslope, below the hill top

And we did this for more than fifteen years, when conditions allowed for it – and in that time I probably had fifteen hundred or more successful such landings. And then one day I didn’t: I lost control of the glider in a strong piece of turbulence just before landing flare, and I crashed hard.
That was not my crash landing, but that’s about what mine would have looked like. For a brief moment after impact I thought I might be dead. For a moment longer, I thought I might be paralyzed. In the end, I got away with a sprained ankle and a moderate case of whiplash.
The accident had happened just prior to a family vacation abroad, and so I had three weeks to think about it while we bounced around on the rutted dirt roads of East Africa and I tried to hold my head still to mitigate the pain in my neck.

As I tried to figure out what had gone wrong, I went over every aspect of the approach. I was flying an intermediate model glider - not very demanding of pilot skill. The conditions were typical for summer – thermals and active air, but not particularly strong or turbulent. I had done hundreds of landings in stronger conditions, on more demanding gliders. At the same time, I was neither sloppy nor complacent in my approach. Though I was relaxed, I was also focused. My intent was simply to fly a perfect approach, and I knew exactly where I wanted to be at every point during the approach – position, heading, altitude and airspeed. On that day I executed the approach exactly as I wanted to.

Finally, after a lengthy analysis during which I could uncover no mistakes in my execution, I was left with only one conclusion. Given the result, which with a little less luck could easily have been a serious injury or death, that landing attempt had been inherently, significantly dangerous.

Consequently, it followed that the error I had made was in the decision to attempt the landing. That decision had been a bad decision, proven so by what had happened, and by what, with a little less luck, could have happened. And given that most of the top landings I had done over the previous fifteen years had involved gliders and conditions that were as demanding or more demanding, that meant that every one of those other decisions had also been bad decisions. And, they had been bad decisions in spite of the fact that no bad outcomes had occurred as a result of any of them. And that’s when the thought occurred to me that I was understanding something for the first time.

My talk today is based on two fundamental ideas: First that the primary determinant of safety is the quality of decision making, and second, that safety in flight test requires that when a negative outcome has potentially catastrophic results, the go decision must be made only when a very high degree of confidence in a positive outcome exists.

Now when this protocol is followed, it is a natural result that a very high percentage of outcomes will be positive, and it is therefore also likely that we will see a long string of consecutive positive outcomes. When this happens, a feedback loop is created, in which each go decision is validated by the positive outcome as having been a good decision, and therefore is reinforced and becomes more likely to be made again in the future under similar circumstances.
But it goes beyond that.

Often, the “lesson learned” from a string of successful outcomes, even if unconsciously, is that our evaluative criteria for the decisions were unnecessarily conservative – not only were we correct to make the go decisions, but we had margin enough that we could have made them even under somewhat less favorable conditions. This can lead to a gradual lowering of our standards.
I want to talk about Murphy’s Law, or at least the statement of Murphy’s law that has gone into the popular lexicon – “Anything That Can Go Wrong Will Go Wrong.” That’s probably not what Murphy said, but that’s how we quote him today. And when we quote Murphy in this way, it is likely that we are doing it mostly for comic or ironic effect. Because to be clear, nobody really believes this statement of Murphy’s Law. After all, almost anything could go wrong, and if it were really true that anything that could go wrong would go wrong, then most everything would go wrong. And most things don’t.

The origin of Murphy’s law is generally attributed to a statement allegedly made by Capt. Edward A. Murphy, a development engineer from Wright Field Aircraft Lab working on Air Force Project MX981 at Edwards Air Force Base in 1949. Commonly known as the rocket sled tests, the project was designed to see how much sudden deceleration a person could withstand in a crash. Murphy had brought a G force sensor, consisting of a bridge of four strain gauges. The gauges could be wired together in more than one way, and in one of those possible ways would cancel each other out and produce a zero reading. After just such a result, Murphy allegedly remarked in frustration, "If there is any way to do it wrong, he will" – referring to the technician who had wired the gauges. This has since been generalized to the now common statement of Murphy’s Law – Anything That Can Go Wrong Will Go Wrong.1,2 This basic observation can be found at least 70 years before Murphy’s statement, and likely goes back much further than that, but today it is almost universally attributed to Murphy.3 The statement is obviously false, since almost anything could go wrong and most things don’t, but it is only false because the common phrasing omits one important word and the related concept. What the statement of the law should be is, “anything that can go wrong, will go wrong, eventually.” That is to say, given enough trials, if something can go wrong, it eventually will. The value of Murphy’s observation is not in the relatively pessimistic implication of the Law as it is quoted today, but rather in the more meaningful way in which it informs design – both the design of things, and the design of procedures. What Murphy is really saying is that if we can design out the possibility of doing it wrong, we should. Our lives today are full of examples of this – the polarized electrical plug being just one.

A tragic illustration of Murphy’s Law was the series of crashes in the F-86’s in which a bolt on the aileron cylinder was installed upside down causing the ailerons to lock up under certain loading conditions. The incorrect bolt installations were reportedly traced to one older worker on the assembly line, who ignored the instructions regarding the orientation of the bolt because he felt he knew better how the bolt should be installed.4

But if Murphy’s law were true, if it were really true that anything that could go wrong would go wrong, then a corollary of Murphy’s law would also be true: “Anything That Did Not Go Wrong, could not have gone wrong.”
And, interestingly enough, even though we don't believe Murphy's Law, we are surprisingly subject to being seduced by this corollary of Murphy's Law - that is, if nothing DID go wrong, then nothing COULD HAVE gone wrong. In other words, we take a positive outcome as evidence that a positive outcome was inevitable.

This is what happened with our decision to use top landings to improve efficiency. We had a long string of successes, each one of which reinforced our perception that success was an outcome we could rely on. And here's the problem with that perception - our data set is too small, at least in the beginning of any such process.

It's important to an understanding of the power of Murphy's corollary to note that we had many indications during those fifteen years of successful landings that the decision to use this technique was of questionable safety. There were a couple of reasons we chose to land on the upslope of the side of the hill, below the top. One reason was that there was inevitably some turbulence, either rotor or convective or both, as we passed through an altitude of about twenty to fifty feet agl. Being weight shift controlled aircraft with relatively large wingspans and light wing loading (33 foot span, 1.1 to 2.0 lbs per square foot) hang gliders suffer from a progressive loss of control authority and response at lower speeds, especially as one approaches stall speed while slowing down. Approaching from below the top allowed us to approach at high speed for better control, and landing on the upslope greatly reduce the duration of time spent at low speed, reducing the probability of a loss of control. Typically, we would pass through the area of turbulence while still flying fast, then enter a region of wind gradient, where airspeed would drop dramatically as we rounded out to follow the slope of the hill, and then almost immediately flared for landing. Many times, while passing through the region of turbulence and then entering the gradient, I remember thinking, “this may not end well.” And yet, with each successful landing, my next thought was, “well that was fine, no problem – what could go wrong?”

I had even observed other pilots having significant crash landings, sometimes even damaging gliders, yet I was able to rationalize those as having been due to poor technique – a poorly executed approach path, or poor control of airspeed on approach. It was only when it happened to me, following what I deemed to be a perfectly executed approach, that I was able to realize that my doubts about the safety had in fact been correct, and the suppression of those doubts had been Murphy’s corollary in action.

Let's imagine we're making "Go" decisions when we have a 99 per cent probability of success. Chances are, we'll see a long string of successful outcomes. And yet, after only 18 such decisions, our overall cumulative probability of success has fallen to 83%.

Raise the percentage probability of an outcome to the power of the number of trials, to compute the overall percentage probability of that outcome for all events in the series.
In other words, after 18 events, our cumulative probability of failure has risen to 17%, or one in six - the same level as in our game of Russian roulette.

\[ 1 - .83 = .17 = 1/6 \]
I don’t know how many of you took me up on my offer of ten million dollars with one chance in six of a catastrophic outcome, but my guess is not many.

And finally, if all of the foregoing were not enough, there is one very significant additional factor that comes into play – the inevitable pressure on us to make the go decision.

This pressure can take many forms, and have many sources, an impending deadline, a project that is behind schedule or over budget or below expectations, or simply our own personal ego, but it is almost always present in some form. The result is that we have a strong incentive to make the go decision, even when we have some level of doubt about it. And beyond that, we have a strengthened tendency to validate the go decision as having been a good one when the outcome is successful, because now the pressure has been relieved.
Conversely, if we do decide to make the no go decision, then the pressure to make the go decision the next time typically only increases, as we are now farther behind on whatever schedule, or budget, or whatever other factors may have been creating the pressure to go in the first place.
I have, over time, come to the conclusion that the pressure to go may be the single most difficult factor to overcome in the pursuit of safety. I have observed in my own decision making that I am far from immune to it, despite the extent of the analysis I have done into its potential for negatively impacting the decision making process in relation to safety.

Ok. So twenty years ago I had a hang gliding accident that led me to start thinking about how we make decisions that impact our safety. Three years after that, I wrote these thoughts out in some detail, in much the same form as I’ve talked about them today, in an article for the national Hang Gliding Magazine. And eventually I began to ask myself whether any of these ideas might have some broader relevance. And you might be asking the same question at this point - how does any of this apply to experimental flight test in the field of aerospace?

I will suggest three possible examples:

1. Apollo 1

We’d had a history of success with a pressurized pure oxygen environment in manned space capsules, until Apollo One. Had we thought sufficiently critically about it, we might have concluded that a complex electrical system in a significantly pressurized pure oxygen environment was a dangerous condition. We had even had previous fires in such environments, and we might have drawn on those as evidence of such danger.

But, in the words of Frank Borman, the astronaut’s representative on the Apollo 204 Review Board, in testimony before the House Subcommittee on NASA Oversight,
“I don’t believe that any of us recognized that the test conditions for this test were hazardous.”⁵ “We did not think, and this is a failing on my part and on everyone associated with us; we did not recognize the fact that we had the three essentials, an ignition source, extensive fuel and, of course, we knew we had the oxygen.”⁶

In the Summary of its report, the Committee on Aeronautical and Space Sciences of the United States Senate had this to say:

“It is clear from the Board’s report and the testimony before the committee that this kind of accident was completely unexpected despite the amount of documentation of fire hazards in pure oxygen environments. The committee can only conclude that NASA’s long history of successes in testing and launching space vehicles with pure oxygen environments at 16.7 p.s.i. and lower pressures led to overconfidence and complacency.”⁷

2. Space Shuttle Challenger

We had successfully launched a number of space shuttles with O-ring erosion and hot gas blow by in the joints of the solid rocket boosters adjacent to the external tank. With sufficiently critical analysis, and better internal communication, we might have concluded that this was a dangerous situation before we had one blow up. Instead, influenced by our repeated successes, we actually expanded both the amount of blow by that was acceptable and the low temperature range at which we were willing to launch, increasing the danger of a catastrophic failure.
Quoting from the Report of the Presidential Commission on the Space Shuttle Challenger Accident:

“The Commission concluded that there was a serious flaw in the decision making process leading up to the launch of flight 51-L. A well-structured and managed system emphasizing safety would have flagged the rising doubts about the Solid Rocket Booster joint seal.”

“Neither Thiokol nor NASA expected the rubber O-rings sealing the joints to be touched by hot gases of motor ignition, much less to be partially burned. However, as tests and then flights confirmed damage to the sealing rings, the reaction by both NASA and Thiokol was to increase the amount of damage considered "acceptable." At no time did management either recommend a redesign of the joint or call for the Shuttle's grounding until the problem was solved.”

“NASA and Thiokol accepted escalating risk apparently because they ‘got away with it last time.’ As Commissioner Feynman observed, the decision making was: ‘a kind of Russian roulette. ... (The Shuttle) flies (with O-ring erosion) and nothing happens. Then it is suggested, therefore, that the risk is no longer so high for the next flights. We can lower our standards a little bit because we got away with it last time. ... You got away with it, but it shouldn't be done over and over again like that.’

We had successfully re-entered and landed a number of shuttles with damaged or missing heat shield tiles that had been struck by foam shedding from the external tank. We might have concluded that the damage to or loss of those tiles created an unacceptable risk on re-entry.
Instead, we continued to launch shuttles with the foam shedding from the tank and damaging the tiles, until we lost Columbia.

Quoting from the Report Of The Columbia Accident Investigation Board:

“The shedding of External Tank foam – the physical cause of the Columbia accident – had a long history. This raises an obvious question: Why did NASA continue flying the Shuttle with a known problem that violated design requirements? It would seem that the longer the Shuttle Program allowed debris to continue striking the Orbiters, the more opportunity existed to detect the serious threat it posed. But this is not what happened”

“With each successful landing, it appears that NASA engineers and managers increasingly regarded the foam-shedding as inevitable, and as either unlikely to jeopardize safety or simply an acceptable risk.”

1. Apollo 1 - 1967
2. Space Shuttle Challenger - 1986
3. Space Shuttle Columbia - 2003

In each of these three cases, the post-accident investigation found that a problem had existed that should have been recognized and corrected, and was not.

I’ve offered these incidents as evidence of relevance to the Society of the ideas I’ve talked about today. I’ve quoted from several accident reports and in so doing, I have clearly demonstrated that nothing I have said here today is either new or original, even though these ideas seemed both new and original when they first occurred to me, as they came out of my analysis of my own accident, and as I had not read any of these NASA reports at that time.
What may be somewhat original, or at least worthy of consideration, in what I’ve talked about today, are some possible insights into why these problems seem so intractable. Why, after these factors have been repeatedly identified in formal accident investigation reports, do they then show up again in subsequent accidents? Why did we have Challenger after Apollo One, and why did we have Columbia after Challenger?

The Report Of The Columbia Accident Investigation Board, among other recommendations, called for a “new culture” at NASA.¹³

I will suggest that as decision makers we can become victims not merely of a culture, but of a deeper and more persistent human psychology.

---

**The Challenge**

- Deceptive Mathematics Of Probability
- Self-Deceptive Psychology Of Evaluating Our Own Decisions
- Inevitable Pressure To Make The “Go” Decision

---

I think that as decision makers we face an unexpectedly difficult challenge arising out of a powerful combination of the inherently deceptive mathematics of probability, the potentially self-deceptive psychology involved in evaluating the quality of our own decisions, and the inevitable significant pressure we experience to make a go decision even when we have doubts about the safety of that decision.

So, if this is a problem, then what is the answer?

I won’t suggest that I have an answer, but I will offer two strategies that I have employed, and that I have found to be helpful.
The first is simply to be aware of the problem; to be continually aware of our inherent tendency to succumb to Murphy’s Corollary, and to the pressure to go forward under less than adequately favorable circumstances.

The second is to expand the scope of source material for our analysis into lessons learned.

We tend to share lessons that are learned from some type of flight test failure. And we analyze these failures, and share these lessons in the hope and belief that doing so will help us avoid repeating the mistakes of the past. And this is of course completely appropriate and worthwhile.

At the outset, I asked if this process might be missing something. This is what I think we may be missing - that there may also be techniques available for identifying and avoiding mistakes of the future – mistakes we don’t know anything about because we’ve not yet seen them.

This might be achieved through a careful and critical pre and post-test analysis of our successes – an analysis aimed at identifying successes which may be wholly, or partially, the result of good fortune as opposed to sound decision making.

One method for doing this is to establish beforehand a quantitative metric that relates to safety, and then to set a target for performance. The target can be set so as to allow for a significant margin of safety.
Then by measuring actual performance against our target, we may identify apparently successful outcomes that did not meet our target performance, and therefore should not, perhaps, be considered successes.

This could have been done during the Shuttle program in relation to the O-ring seals, and it would have triggered a stop to the program until the problem was fixed. We might guess why that wasn’t done – the program was under performing and the pressure to go, which can exist at many levels within a complex organization, was likely significant.

The existence and/or perception of pressure was referenced a number of times in the Rogers report. The commission noted for example: “The nation’s reliance on the Shuttle as its principal space launch capability created a relentless pressure on NASA to increase the flight rate.” Roger Boisjoly, a structures engineer at Thiokol, who strongly opposed the launch of Challenger under the temperature conditions at the time, testified before the commission, “I felt personally that management was under a lot of pressure to launch and that they made a very tough decision, but I didn’t agree with it.” Interestingly, it does not really matter whether pressure is “real” or merely perceived – it has the same effect. In Moonshot, Alan Shepard and Deke Slayton wrote, regarding Skip Chauvin’s decision to decline an engineer’s suggestion to cancel the plugs out test due to multiple problems, “Time was more important. It was becoming damned critical with all the pressures mounting to get this ship on its way into orbit.”
Ironically, however, it’s not unlikely that the program lost more time in fixing the problem after the catastrophe than would have been lost to a proactive approach to fixing it ahead of time.

A rigorous analysis of successful outcomes would constitute an expansion of our source material for lessons learned. If such a process allows us to uncover some decisions or procedures that may have been in some way flawed, we may expand the degree to which our lessons learned enhance flight test safety.

Acknowledgments

The author gratefully acknowledges the assistance of Colin A. Fries and the NASA HQ History Program Office in gaining access to original source material used for this paper.
Notes:


6. Ibid., 87


9. Idid., Chapter VI – Early Design

10. Ibid., Chapter VI – Findings, Item 3.


12. Ibid.

13. Ibid. Chapter 11 – Recommendations
14. Rogers, et al, op. cit., Chapter IX, Section VII – Flight Rate

15. Ibid., Chapter V

References:


Copyright © 2015 by Michael W. Meier – All Rights Reserved