CHAPTER 11

Learning to Drop Your Familiar Tools

JAKE, THE ATHLETIC-LOOKING sandy blond, speaks first. He wants to race the car. “What if everybody just agrees?” he asks. “I say, race this thang.”

It was early afternoon in fall, and Jake and six of his second-year Harvard Business School classmates found a shady spot where they could eat their lunches and talk.* Their professor had given them three pages containing one of the most famous business school case studies ever created, known as Carter Racing. The crux is whether the fictional Carter Racing team’s car should compete in the biggest race of the season, which begins in one hour.

The argument in favor of racing: thanks to a custom turbocharger, Carter Racing has placed in the money (top five) in twelve of twenty-four races. That success secured an oil company sponsorship, and a trial sponsorship from prestigious (and also fictional) Goodstone Tire. Carter Racing won the last race, its fourth win of the season. Today’s race will be on national TV, and if Carter Racing finishes in the top five, it will likely draw a $2 million sponsorship from Goodstone. If Carter Racing chooses not to race and withdraws, it would lose part of its entry fee and have to pay back some sponsor money. The team would end a stellar season $80,000 in the hole, and may never get another shot this big. Racing seems like a no-brainer.

The argument against racing: in seven of twenty-four races, the engine failed, each time damaging the car. In the last two races, the mechanics used a new engine-prep procedure and had no trouble, but they aren’t sure what caused the problem before. If the engine fails on national TV, the team will lose the oil sponsorship, kiss Goodstone goodbye, and go back to square one, or perhaps out of business. So: race, or don’t race?

The group begins with a vote. Three students vote to race, four to sit it out. Now the debate begins.

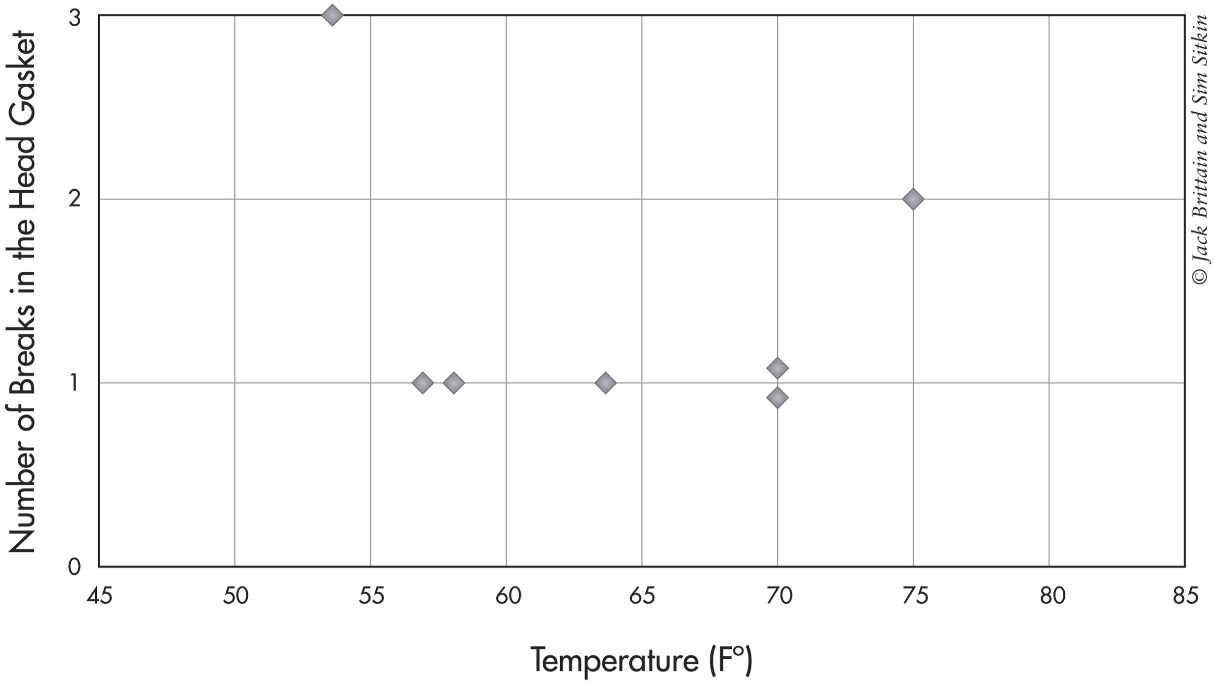
Even with the engine failures, Jake says, the team has a 50 percent chance of its biggest triumph. The upside of the Goodstone sponsorship is much more money than the team stands to lose if the engine fails and the existing sponsors walk. If Carter Racing withdraws, an excellent season ends with debt, “which, as we all know, is not a sustainable business model.”

“I just don’t think they can afford not to race,” Justin says.

Alexander agrees, and addresses the dissenters: “What’s going to change going forward to convince you that now you’re ready?” he asks.

Mei, wearing a Harvard hoodie and sitting across the circle, has a calculation to share. “To me, the risk of not racing is about one-third of the downside of [another engine failure],” she says. She adds that she’s focusing on loss mitigation, and does not want to race.

The case study says that at the last minute, the team owner, BJ Carter, called his mechanics. Pat, the engine mechanic, dropped out of high school and has no sophisticated engineering training, but he has a decade of race experience. Temperature could be the issue, he suggested. When the turbocharger warms up on a cool day, engine components might expand at different rates and set up failure of the head gasket, a metal seal in the engine. Pat admitted that each engine failure looked different, but all seven had breaks in the head gasket. (Two of the engine failures had multiple breaks in the gasket.) He didn’t know what was going on, but couldn’t think of anything else on short notice. He was still hyped to race, and jubilant about the new Goodstone uniforms. At 40 degrees, it is the coldest race day of the season. Robin, the chief mechanic, endorsed Pat’s idea to look at the temperature data. He plotted it on a graph, but saw no correlation:



Dmitry, his black hair flopped to one side, is firmly against racing. He agrees that there is no apparent linear relationship between gasket failure and temperature; three gasket breaks occurred on the coolest race day (53 degrees), and two on one of the hottest days (75 degrees). But what if there is an optimal range for the engine, not too cold and not too hot? “If the failures are random, the probability that you both finish and get in the top five is 50 percent,” Dmitry says. “But if it’s not random, the probability is lower. This day is a very, very low temperature that they haven’t experienced before. We don’t know if there’s a correlation with temperature, but if there is, it’s like a sure thing that it fails.”

Julia thinks mechanic Pat’s temperature idea is “nonsense,” but like Dmitry views the engine problem as a black box that does not give the team any information to calculate probability for today’s race. She acknowledges that she’s being risk averse, and would personally never get involved in car racing at all.

Except for Dmitry, the group agrees that there is “zero correlation at all,” as Alexander puts it, between temperature and engine failure. “Am I the only one?” Dmitry asks, to a few giggles.

Jake is particularly unimpressed with engine mechanic Pat’s reasoning. “I think Pat’s a really good mechanic,” he says. “I don’t think he’s a really good root cause analysis engineer, and those are two very different things." Jake thinks Pat is falling prey to a well-known cognitive bias, overemphasizing the importance of a single, dramatic memory—the three gasket breaks on a cool day. "We don't even have the information to understand this graph," Jake says. "There's twenty-four races, right? How many of those were around 53 degrees and didn't break? I don't mean to attack your point," he says to Dmitry, smiling and giving him a friendly tap on the hand.

Everyone agrees it would be nice to have temperature data from the races with no engine problems, but that they're stuck with what they have. Justin speaks for the entire pro-race side when he says, "I just think you've gotta race, because that's what you're in this business to do."

It seems that the group will finish where they started, voting not to race, until Mei takes another look at her calculations. "I've actually changed my mind," she announces. "I'm voting for yes, race." Comparing the potential financial upside and downside, Mei calculated that Carter Racing needs just a 26 percent chance of finishing in the top five—half their current rate—to make racing a smart bet. Even if the cool temperature changes the odds, "it won't decrease it to 26 percent, so we are still safe." She thinks Dmitry's read of the data is biased; Carter Racing has competed at temperatures from 53 to 82 degrees, with four engine failures below 65 and three above. Dmitry is giving too much credence, Mei says, to the 53-degree data point because it involved three gasket breaks. It's still just one engine failure.

Jake jumps in and says that group members are seeing whatever they want in the temperature chart, so "maybe we table that debate." He likes Mei's expected value argument. "I think that's one concrete thing we can go with, in terms of it's always good to base things on math. . . . If you told me to flip a coin, and if I lose the flip I lose $100 but if I win I get $200, I flip that coin every time." He reminds the group that Carter Racing used a new engine-prep procedure for the last two races, with no problems. "That's a small data point," he says, "but at least it's in the right direction for my argument."

Mei turns to Dmitry. "What is the temperature you feel comfortable to race?" she asks. "We have two engine failures at 70, one at 63, and one at 53. There's no temperature that's safe for us."

Dmitry wants to set limits at exactly the temperatures they have already experienced. Something is not functioning as expected, so anything outside that temperature range is unknown territory. He knows his recommendation comes off as extremely arbitrary.

The group moves to a final tally. With Mei's conversion, it's four to three, they're racing. The students continue to chat as they stuff the case study papers into their backpacks and messenger bags.

Martina quickly reads aloud a part of the case study where team owner BJ Carter asked his chief mechanic, Robin, for his opinion. "The drivers have their lives on the line, I have a career that hangs on every race, and you have every dime tied up in the business," Robin told him. Nobody ever won a race sitting in the pits, he reminded his boss.

Martina has one last question. "This is just about money, right? We're not going to kill anyone if we race, are we?"

A few of the group members look around and laugh, and then they go their separate ways.

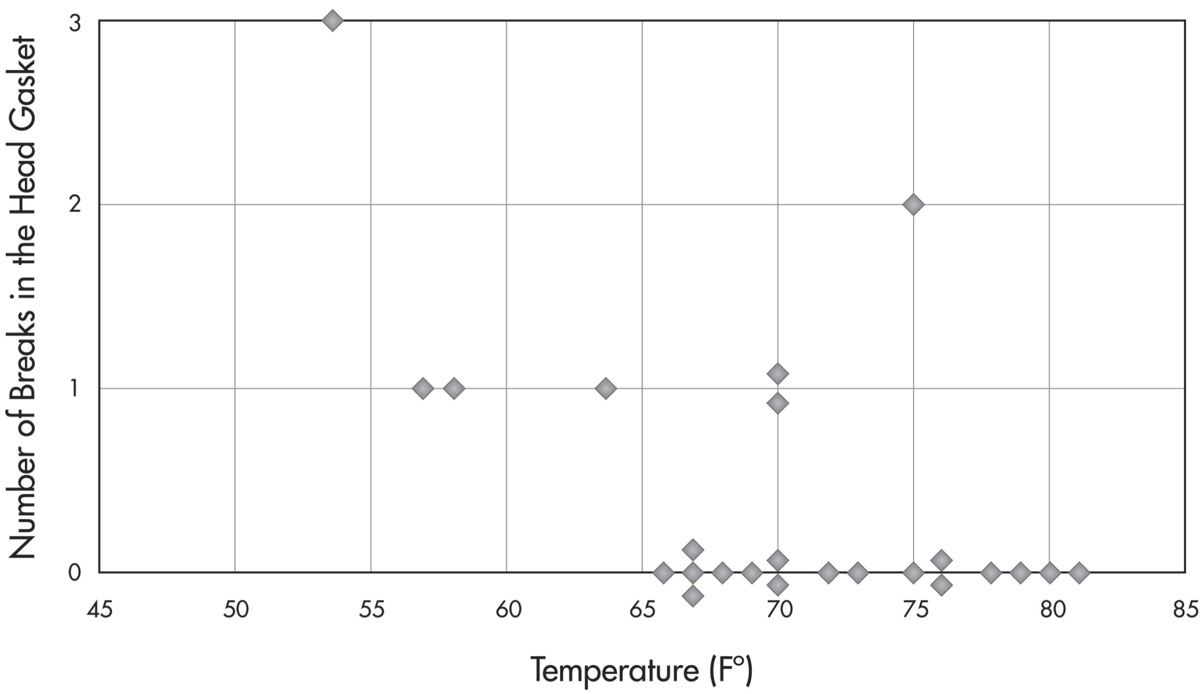
• • •

When the students arrive in class the next day, they learn that most student groups around the world who have ever been assigned the Carter Racing case chose to race. The professor goes around the room, interrogating their logic for racing or withdrawing.

Teams that decided to race discuss their probability estimates and decision trees. Students are split on whether mid-race engine failure will endanger the driver. A majority of students think the temperature data is a red herring. Heads nod when one woman says, "If we want to make something of ourselves in the business of racing, this is the kind of risk we need to take." Her team was unanimous, 7–0, for race.

Dmitry objects, and the professor grills him ruthlessly. Dmitry contends that every probability decision tree that every group posits is irrelevant if you drop the assumption that engine failures are randomly distributed. He adds that the data are particularly ambiguous because for some reason the chief mechanic didn't plot the race temperatures when the engine didn't fail.

“Okay, so, Dmitry, here comes a quantitative question,” the professor says. “How many times did I say yesterday if you want additional information let me know?” Muffled gasps spread across the room. “Four times,” the professor answers himself. “Four times I said if you want additional information let me know.” Not one student asked for the missing data. The professor puts up a new graph, with every race plotted. It looks something like this:



Every single race below 65 degrees had an engine failure. The professor then labels every race either a fail or not fail, and with that binary division runs a simple statistical analysis, familiar to the students, known as a logistic regression. He informs the students that there is a 99.4 percent probability of engine failure at 40 degrees. “Do we have any remaining fans of racing?” he asks. And now he has another surprise.

The temperature and engine failure data are taken exactly from NASA’s tragic decision to launch the space shuttle *Challenger*, with the details placed in the context of racing rather than space exploration. Jake’s face goes blank. Rather than a broken gasket, *Challenger* had failed O-rings—the rubber strips that sealed joints along the outer wall of the missile-like rocket boosters that propelled the shuttle. Cool temperatures caused O-ring rubber to harden, making them less effective seals.

The characters in the case study are loosely based on managers and engineers at NASA and its rocket-booster contractor, Morton Thiokol, on an emergency conference call the night before the *Challenger* launch. Weather reports on January 27, 1986, predicted unusually cool Florida weather for launch. After the conference call, NASA and Thiokol gave the okay to proceed. On January 28, O-rings failed to properly seal a joint in the wall of a rocket booster. Burning gas shot right through the joint to the outside, and *Challenger* exploded seventy-three seconds into its mission. All seven crew members were killed.

The Carter Racing case study worked exquisitely. It was eerie how precisely the students filled the shoes of the engineers on the emergency conference call who gave the green light for launch. The professor unfurled the lesson masterfully.

"Like all of you, nobody [at NASA or Thiokol] asked for the seventeen data points for which there had been no problems," he explains. "Obviously that data existed, and they were having a discussion like we had. If I was in your situation I would probably say, 'But in a classroom the teacher typically gives us material we're supposed to have.' But it's often the case in group meetings where the person who made the PowerPoint slides puts data in front of you, and we often just use the data people put in front of us. I would argue we don't do a good job of saying, 'Is this the data that we want to make the decision we need to make?'"

The presidential commission that investigated the *Challenger* accident concluded that simply including the nonfailure flights would have revealed the correlation between O-ring damage and temperature. A University of Chicago professor of organizational psychology wrote that the missed data was such a rudimentary mistake that it came down to "a professional weakness shared by all participants" on the conference call. "Arguments against launching at cold temperatures could have been quantified, but were not quantified." The engineers were poorly educated, he declared.

Sociologist Diane Vaughan's book *The Challenger Launch Decision* came to be regarded by NASA as the definitive causal account of the tragedy. "More stunning is the observation that they *did* have the pertinent data," it reads. "There were charts [that several Thiokol engineers who wanted to postpone launch] did not imagine and did not construct that, if created, would have provided the quantitative correlational data required to sustain their position."

Business professors around the world have been teaching Carter Racing for thirty years because it provides a stark lesson in the danger of reaching conclusions from incomplete data, and the folly of relying only on what is in front of you.

And now for one last surprise. They all got it wrong. The *Challenger* decision was not a failure of quantitative analysis. NASA's real mistake was to rely on quantitative analysis too much.

• • •

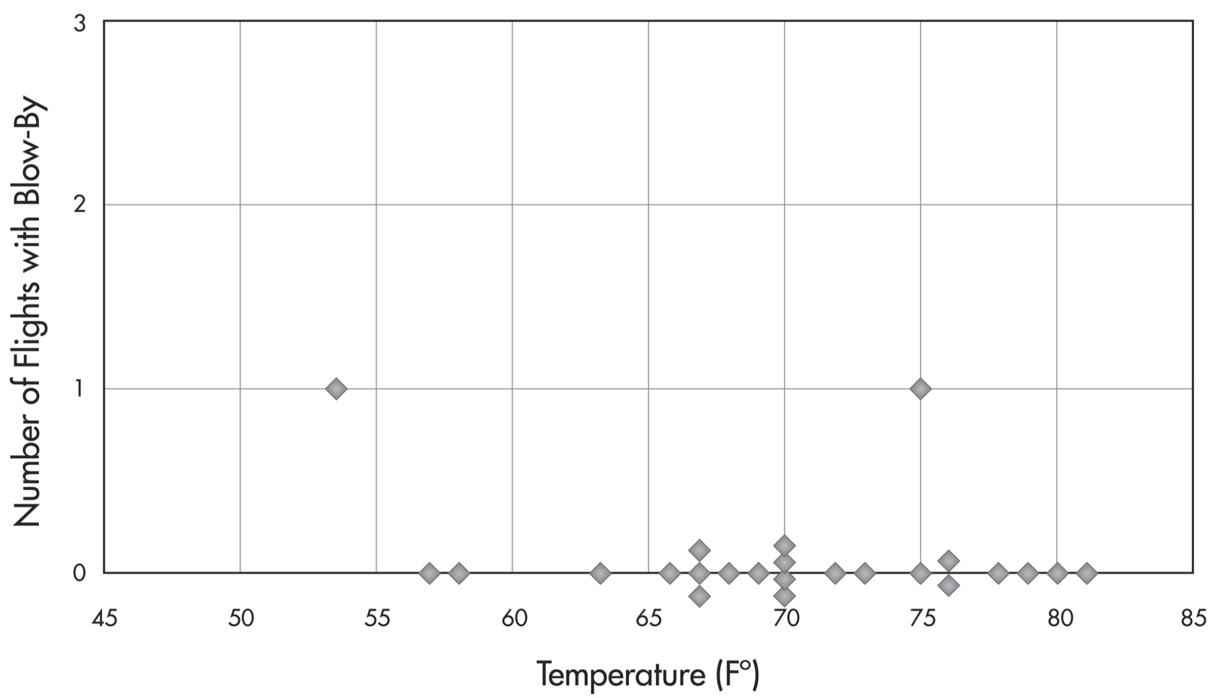
Before ignition, *Challenger*'s O-rings sat squashed in the joints that connected vertical sections of the booster. At ignition, burning gas came shooting down the booster. The metal walls that connected to form a joint pulled apart for a split second, at which point the rubber O-rings immediately expanded to fill the space and keep the joint sealed. When the O-rings got cold, the rubber hardened and could not expand as quickly. The colder the O-ring, the longer the fraction of a second when the joint was not sealed and burning gas could shoot right through the booster wall. Even so, temperature usually did not matter; the O-rings were protected by a special insulating putty meant to block burning gas from reaching them in the first place. On the seventeen flights with no O-ring problems—akin to the seventeen Carter Racing races with no engine problems—the putty worked perfectly. Those flights provided no information whatsoever about how O-rings might fail, no matter the temperature, because the burning gas could not even get to the O-rings to cause a problem. Sometimes, however, small holes formed in the putty when the joints were assembled. On the seven flights that had O-ring issues, burning gas pushed through the holes in the protective putty and reached the O-rings. Only those seven data points were relevant to how the O-rings could be damaged or fail.

And on those seven shuttle flights—unlike gasket breaks in Carter Racing, which was the same problem every time—the O-ring issues came in two different varieties. The first: erosion. On five flights, burning gas that came shooting down the booster at ignition hit the O-rings and eroded the rubber surface. This was not a life-or-death condition. There was more than enough rubber for the O-ring to do its job. And erosion had nothing at all to do with temperature.

The second variety: blow-by. If the rubber ring did not expand instantly to fully seal the joint at ignition, burning gas “blew by” and could potentially shoot right through the booster wall. Blow-by *was* a life-or-death condition and, engineers would later learn, dramatically worsened when cool temperatures hardened the O-ring rubber. Two pre-*Challenger* flights had blow-by, but still returned home safely.

Thiokol engineers who opposed the launch on the emergency prelaunch conference call did not really have twenty-four relevant data points on O-ring failure to work with, as the Carter Racing study indicates. They did not even have seven, like the Harvard students. They had two.

Now what does the chart tell you?



Ironically, Allan McDonald, then director of the rocket-booster project at Morton Thiokol, told me, “Looking only at the relevant data points supported NASA’s [prelaunch] position that it was inconclusive.” There was no 99.4 percent certainty that was missed. The engineers were not poorly educated.

There was other important information the Thiokol engineers presented that could have helped NASA avert disaster. But it was not quantitative, so NASA managers did not accept it. The Carter Racing study teaches that the answer was available, if only engineers looked at the right numbers. In reality, the right numbers did not contain an answer at all. The *Challenger* decision was truly ambiguous. It was a wicked problem, rife with uncertainty, and outside of previous experience, where demanding more data actually became the problem itself.

• • •

The infamous emergency conference call convened thirty-four engineers —every manager was also an engineer—in three locations. Thiokol engineer Roger Boisjoly had personally inspected the joints after both flights with blow-by, and presented photographs from each. Following the 75-degree flight, he found a very thin streak of light gray soot beyond an O-ring in the joint, from a tiny amount of gas that had blown by before the O-ring sealed. It was nowhere close to a catastrophic problem. After the 53-degree flight, he found jet-black soot fanned out across a large swath of the joint. A lot of burning gas had blown past that time. In Boisjoly's opinion, the reason the 53-degree launch looked so much worse was that cool conditions had hardened the O-rings and made them slow to expand and seal at ignition. He was right, but he did not have the data to prove it. "I was asked to quantify my concerns, and I said I couldn't," Boisjoly later testified. "I had no data to quantify it, but I did say I knew that it was away from goodness."

Thanks to an extraordinarily strong technical culture, NASA had developed quantitatively rigorous "flight readiness reviews." They were productively adversarial, like superforecasting team discussions. Managers grilled engineers and forced them to produce data to back up their assertions. The process had worked remarkably. The space shuttle was the most complex machine ever built, and all twenty-four flights had returned safely. But on the emergency conference call, that same quantitative culture led them astray.

On their engineers' advice, McDonald and two Thiokol VPs on the call initially supported a no-launch decision. The *Challenger* had already been cleared, so this was an eleventh-hour reversal. When NASA officials asked Thiokol engineers exactly what temperature range was safe for flight, they recommended setting a limit at 53 degrees, the lower bound of previous experience.

NASA manager Larry Mulloy was flabbergasted. He thought the shuttle was supposed to be cleared to launch from 31 to 99 degrees. A last-minute 53-degree limit was setting an entirely new technical criteria for launches. It had never been discussed, was not backed by quantitative data, and meant that suddenly winter was off-limits for space exploration. Mulloy found it frustrating; he later called it "dumb."

How had the engineers arrived at that number? "They said because they had flown at 53 degrees before," a NASA manager reflected, "which is no reason to me. That's tradition rather than technology." Boisjoly was asked again for data to support his claim, "and I said I have none other than what is being presented."

With the conference call at an impasse, a Thiokol VP asked for a five-minute "offline caucus," during which Thiokol concluded that they had no more data to provide. They returned to the call a half hour later with a new decision: proceed with launch. Their official document read, "temperature data not conclusive on predicting primary O-ring blow-by."

When conference call participants from NASA and Thiokol later spoke with investigators and gave interviews, they repeatedly brought up the "weak engineering position," as one put it. Their statements comprised a repetitive chorus: "Unable to quantify"; "supporting data was subjective"; "hadn't done a good technical job"; "just didn't have enough conclusive data." NASA was, after all, the agency that hung a framed quote in the Mission Evaluation Room: "In God We Trust, All Others Bring Data."

"The engineers' concerns for the most part were just based on a few photographs they took of joints they pulled apart that had soot trapped in there," McDonald told me. "One was at a cool temperature, and one was at a rather warm temperature. Roger Boisjoly thought the difference was absolutely telling a story, but it was a qualitative assessment." NASA's Mulloy later argued that he "would've felt naked" taking Thiokol's argument up the chain of command. Without a solid quantitative case, "I couldn't have defended it."

The very tool that had helped make NASA so consistently successful, what Diane Vaughan called "the original technical culture" in the agency's DNA, suddenly worked perversely in a situation where the familiar brand of data did not exist. Reason without numbers was not accepted. In the face of an unfamiliar challenge, NASA managers failed to drop their familiar tools.

• • •

Psychologist and organizational behavior expert Karl Weick noticed something unusual in the deaths of smokejumpers and "hotshot" wilderness firefighters: they held on to their tools, even when ditching equipment would have allowed them to run away from an advancing fire. For Weick, it was emblematic of something larger.

In Montana's 1949 Mann Gulch fire, made famous in Norman Maclean's *Young Men and Fire*, smokejumpers parachuted in expecting to face a "ten o'clock fire," meaning they would have it contained by 10 a.m. the next morning. Until the fire jumped across the gulch from one forested hill slope to the steep slope where the firefighters were, and chased them uphill through dry grass at eleven feet per second. Crew foreman Wagner Dodge yelled at the men to drop their tools. Two did so immediately and sprinted over the ridge to safety. Others ran with their tools and were caught by the flames. One firefighter stopped fleeing and sat down, exhausted, never having removed his heavy pack. Thirteen firefighters died. The Mann Gulch tragedy led to reforms in safety training, but wildland firefighters continued to lose races with fires when they did not drop their tools.

In 1994, on Colorado's Storm King Mountain, hotshots and smokejumpers faced a Mann Gulch situation when a fire jumped a canyon and erupted through a stand of gambel oak below them. The sound in the canyon was "like a jet during take off," according to a survivor. Fourteen men and women lost the race with a wall of flame. "[Victim] was still wearing his backpack," reads an analysis from the body recovery operation. "Victim has chainsaw handle still in hand." He was just 250 feet from a safe zone. Survivor Quentin Rhoades had already run nine hundred feet uphill, “then realized I still had my saw over my shoulder! I irrationally started looking for a place to put it down where it wouldn’t get burned. . . . I remember thinking I can’t believe I’m putting down my saw.” Two separate analyses conducted for the U.S. Forest Service and the Bureau of Land Management concluded that the crew would have made it out intact had they simply dropped their tools and run from the start.

In four separate fires in the 1990s, twenty-three elite wildland firefighters refused orders to drop their tools and perished beside them. Even when Rhoades eventually dropped his chainsaw, he felt like he was doing something unnatural. Weick found similar phenomena in Navy seamen who ignored orders to remove steel-toed shoes when abandoning a ship, and drowned or punched holes in life rafts; fighter pilots in disabled planes refusing orders to eject; and Karl Wallenda, the world-famous high-wire performer, who fell 120 feet to his death when he teetered and grabbed first at his balance pole rather than the wire beneath him. He momentarily lost the pole while falling, and grabbed it again in the air. “Dropping one’s tools is a proxy for unlearning, for adaptation, for flexibility,” Weick wrote. “It is the very unwillingness of people to drop their tools that turns some of these dramas into tragedies.” For him, firefighters were an example, and a metaphor for what he learned while studying normally reliable organizations that clung to trusty methods, even when they led to bewildering decisions.

Rather than adapting to unfamiliar situations, whether airline accidents or fire tragedies, Weick saw that experienced groups became rigid under pressure and “regress to what they know best.” They behaved like a collective hedgehog, bending an unfamiliar situation to a familiar comfort zone, as if trying to will it to become something they actually had experienced before. For wildland firefighters, their tools are what they know best. “Firefighting tools define the firefighter’s group membership, they are the firefighter’s reason for being deployed in the first place,” Weick wrote. “Given the central role of tools in defining the essence of a firefighter, it is not surprising that dropping one’s tools creates an existential crisis.” As Maclean succinctly put it, “When a firefighter is told to drop his firefighting tools, he is told to forget he is a firefighter.”

Weick explained that wildland firefighters have a firm "can do" culture, and dropping tools was not part of it, because it meant they had lost control. Quentin Rhoades's chainsaw was such a part of his firefighting self that he did not even realize he still had it, any more than he realized he still had his arms. When it became utterly ludicrous to carry the saw further, Rhoades still "could not believe" he was parting with it. He felt naked, just as Larry Mulloy said he would have without a quantitative argument for a last-second launch reversal. At NASA, accepting a qualitative argument was like being told to forget you are an engineer.

When sociologist Diane Vaughan interviewed NASA and Thiokol engineers who had worked on the rocket boosters, she found that NASA's own famous can-do culture manifested as a belief that everything would be fine because "we followed every procedure"; because "the [flight readiness review] process is aggressive and adversarial"; because "we went by the book." NASA's tools were its familiar procedures. The rules had always worked before. But with *Challenger* they were outside their usual bounds, where "can do" should have been swapped for what Weick calls a "make do" culture. They needed to improvise rather than throw out information that did not fit the established rubric.

Roger Boisjoly's unquantifiable argument that the cold weather was "away from goodness" was considered an emotional argument in NASA culture. It was based on interpretation of a photograph. It did not conform to the usual quantitative standards, so it was deemed inadmissible evidence and disregarded. The can-do attitude among the rocket-booster group, Vaughan observed, "was grounded in conformity." After the tragedy, it emerged that other engineers on the teleconference agreed with Boisjoly, but knew they could not muster quantitative arguments, so they remained silent. Their silence was taken as consent. As one engineer who was on the *Challenger* conference call later said, "If I feel like I don't have data to back me up, the boss's opinion is better than mine."

Dropping familiar tools is particularly difficult for experienced professionals who rely on what Weick called overlearned behavior. That is, they have done the same thing in response to the same challenges over and over until the behavior has become so automatic that they no longer even recognize it as a situation-specific tool. Research on aviation accidents, for example, found that “a common pattern was the crew’s decision to continue with their original plan” even when conditions changed dramatically.

When Weick spoke with hotshot Paul Gleason, one of the best wildland firefighters in the world, Gleason told him that he preferred to view his crew leadership not as decision making, but as sensemaking. “If I make a decision, it is a possession, I take pride in it, I tend to defend it and not listen to those who question it,” Gleason explained. “If I make sense, then this is more dynamic and I listen and I can change it.” He employed what Weick called “hunches held lightly.” Gleason gave decisive directions to his crew, but with transparent rationale and the addendum that the plan was ripe for revision as the team collectively made sense of a fire.

On the night of the *Challenger* conference call, following procedure in the face of uncertainty was so paramount that NASA’s Mulloy asked Thiokol to put its final launch recommendation and rationale on paper and sign it. Last-minute sign-off had always been verbal in the past. Thiokol’s Allan McDonald was in the room with Mulloy, and refused. One of McDonald’s bosses in Utah signed and faxed the document instead. Even Mulloy, who had demanded data, must have felt uneasy with the decision, while at the same time feeling protected by NASA’s ultimate tool—its hallowed process. The process culminated with more concern for being able to defend a decision than with using all available information to make the right one. Like the firefighters, NASA managers had merged with their tools. As McDonald said, looking only at the quantitative data actually supported NASA’s stance that there was no link between temperature and failure. NASA’s normal quantitative standard was a dearly held tool, but the wrong one for the job. That night, it should have been dropped.

It is easy to say in retrospect. A group of managers accustomed to dispositive technical information did not have any; engineers felt like they should not speak up without it. Decades later, an astronaut who flew on the space shuttle, both before and after *Challenger,* and then became NASA’s chief of safety and mission assurance, recounted what the “In God We Trust, All Others Bring Data” plaque had meant to him: “Between the lines it suggested that, ‘We’re not interested in your opinion on things. If you have data, we’ll listen, but your opinion is not requested here.’”

Physicist and Nobel laureate Richard Feynman was one of the members of the commission that investigated the *Challenger,* and in one hearing he admonished a NASA manager for repeating that Boisjoly's data did not prove his point. "When you don't have any data," Feynman said, "you have to use reason."

These are, by definition, wicked situations. Wildland firefighters and space shuttle engineers do not have the liberty to train for their most challenging moments by trial and error. A team or organization that is both reliable and flexible, according to Weick, is like a jazz group. There are fundamentals—scales and chords—that every member must overlearn, but those are just tools for sensemaking in a dynamic environment. There are no tools that cannot be dropped, reimagined, or repurposed in order to navigate an unfamiliar challenge. Even the most sacred tools. Even the tools so taken for granted they become invisible. It is, of course, easier said than done. Especially when the tool is the very core of an organization's culture.

• • •

As Captain Tony Lesmes described it, his team at Bagram Air Base in northeast Afghanistan only went to work when someone got really unlucky. Lesmes commanded a team of Air Force pararescue jumpers, PJs for short, a division of Special Operations designed for harrowing rescue missions, like parachuting into enemy territory at night to save downed pilots. Cross a soldier, a paramedic, a rescue diver, a firefighter, a mountain rescue specialist, and a parachutist, and you get a PJ. Their emblem depicts an angel with arms wrapped around the world, and the words "That others may live."

There was no typical day for the PJs at Bagram. One day they were rappelling down a mountain to rescue a soldier who fell into an unmarked well. Another day they were rushing to treat Marines injured in a firefight. PJs could accompany units out on missions, but mostly they stayed on twenty-four-hour alert, waiting for a "9-line," a form (with nine lines) that provided basic information about an active emergency. Like one that came in on an autumn day in 2009. It was a category alpha, traumatic injuries. Within minutes, the team would be airborne.

Intel was sparse. A roadside bomb had exploded in the middle of an Army convoy of armored vehicles. The site was approximately a half hour away by helicopter. There were serious injuries, but it was unclear how many or how serious, and whether the bomb was part of a search and rescue trap, where enemies lie in ambush awaiting the rescue team.

The PJs were used to working with cloudy information, but this was ambiguous even for them. Lesmes knew they would have to bring heavy equipment, like the Jaws of Life and a diamond-tipped saw, because "you can't just cut through an armored vehicle like a car door," he told me. Weight was an obstacle, especially at altitude in the mountains. If the choppers were too heavy, they wouldn't grab enough air to stay aloft. Fuel limitations were a challenge. Space was a bigger one. Each PJ came with gear, and each of the two helicopters only had interior space on the order of a large van. They didn't know how many soldiers were injured badly enough to need evacuation, and how much space they would need for them.

Lesmes was certain of just one thing: he wanted to make sure they saved enough room for potential patients so that they would only have to visit the explosion site once. It would take extra time to treat and load severely wounded soldiers. The more time on site, the more likely the operation would draw enemy attention. The rescue team could end up needing a rescue team.

He was twenty-seven, and the previous year had led a stateside hurricane rescue team. Afghanistan was his first extended deployment, and he was directing a team with older members who had had numerous overseas deployments. As usual, Lesmes brought two team members to the operations center to get information and help him make sense of the situation. "Sometimes other guys are able to get really good questions out that I wouldn't normally think of," he told me. "And you want to share as much information as possible, and there isn't a lot of time." But there was little additional intel. "In Hollywood, a drone flies over the site and you get all the information," Lesmes told me. "But that's Hollywood."

He walked out to the helicopters, where PJs were donning their full battle rattle, as he put it. The situation didn't fit the usual decision trees; he laid out the challenges, and asked the men: How do we solve this?

Just move equipment around to cram more stuff into the helicopters, one team member suggested. Another said they could leave a few PJs with the Army convoy if they needed extra helicopter room for patients. One recommended they evacuate the most serious patients, and if a second trip was needed, move the convoy from the explosion site and meet them somewhere less conspicuous. But the bomb had exploded in the middle of a procession of vehicles, in rugged terrain. Lesmes didn't even know how mobile the convoy would be.

"We weren't coming up with any real solution that would give us an advantage. I wanted a speed advantage, and the ability to leverage the weight and space for wounded soldiers," he told me. "The distance and the timeline and the constraints and the unknown of the enemy all started to add up. I just started feeling like we didn't have the setup to be successful in a worst-case scenario. There wasn't that pattern recognition, it was outside of the normal pattern." In others words, he didn't have the definitive intel he would have liked. Based on the information he had, Lesmes guessed there would be more than three serious injuries but fewer than fifteen. An idea started to form, one that could preserve more space for potential patients. He could put aside a tool he had never dropped in this situation: himself.

Lesmes had never not accompanied his team on a mass-casualty category alpha. He was the site conductor. His role was to keep a broad view of the situation while PJs were "heads down" working furiously to save patients, or their limbs. He helped secure the site; communicated with his guys, the base, and helicopter pilots who were circling waiting to pick up patients and go; he radioed planes for backup if a firefight erupted; he coordinated with officers in the area, frequently from other military branches. Emotional chaos was an explosion site certainty. Soldiers watching their shell-shocked teammates suck on fentanyl lollipops, in danger of bleeding out, are desperate to help, but they must be moved. The site had to be managed. This time, as long as there were not many more injuries than Lesmes guessed, he knew his senior enlisted team member could manage leadership on the ground while administering medical aid. Lesmes could help ready the field hospital for returning patients, and coordinate helicopter pickups from the operations center, adjusting as he listened via radio to his guys on the ground. It was a trade-off, but every option was.

Lesmes went to the team with his "hypothesis," as he called it—his hunch held lightly. "I wanted them to disprove it," he told me. He told them he planned to stay at the base to save room for equipment and patients. The helicopter blades were spinning up, moments ticking away in the so-called golden hour, the critical window for saving a severely injured soldier. He told them to talk quickly, and he would consider everything they had to say. A few were quiet. Several objected. Togetherness was their most basic tool, the one they didn't know could be dropped until someone said to drop it. One of the men said flatly that it was the commanding officer's job to come along, and he should do his job. Another got angry. A third reflexively suggested that Lesmes was afraid. He told Lesmes that when it was his time, it was his time, so they should just do what they always did. Lesmes *was* afraid, but not for his life. "If something bad happens, and the officer is not there," he told me, "think about explaining that to ten families."

I was sitting with him at the World War II Memorial in Washington, D.C., when he said that. He had been stoic, and then he started crying. "The whole construct is built on that training and that familiarity and that cohesion," he said. "I totally understand why some guys were upset. It was breaking our standard operating procedure. I mean, my judgment was questioned. But if I go, we might have to go to the rescue site twice." The objections he got were emotional and philosophical, not tactical. They had changed his mind about a plan before, but not this time. He would stay, and it was time for them to go. The helicopters strained into the air as Lesmes returned to the operations center. "I struggled immensely," he said. "I could see what was going on, and if something bad were to happen, I could literally watch the rescue helicopter go down."

The rescue mission, thankfully, was an unqualified success. PJs treated injuries at the explosion site, and seven wounded soldiers had to be loaded into the helicopters. They were packed in like sardines. Several required amputations at the field hospital, but all survived.

When it was over, the senior enlisted man acknowledged it was the right call. Another PJ did not address it for months, and then only to say that he was taken aback that Lesmes had that much trust in them. The soldier who had gotten angry initially remained angry, for a while. Another Bagram PJ I spoke with said, "If I was in that position, I definitely would have said, 'Yeah, we're all going.' It must have been really hard."

"I don't know, man," Lesmes told me. "Sometimes, I still struggle with that decision. Something could've gone wrong and then it would be a bad decision. Maybe it was luck. None of the options at the time looked very optimal."

As we finished talking, I mentioned Weick’s work about wilderness firefighters clinging to their tools. Under pressure, Weick explained, experienced pros regress to what they know best. I suggested to Lesmes that maybe his PJs were just reacting emotionally, with a reflex for the familiar. There must be times when even the sacrosanct tool of togetherness should be dropped, right? “Yeah, mmm-hmm.” He nodded in agreement. It was, of course, easy for me to say. He paused for a moment. “Yeah,” he said, “but everything is built on that.”

• • •

The *Challenger* managers made mistakes of conformity. They stuck to the usual tools in the face of an unusual challenge. Captain Lesmes dropped a sacred tool, and it worked. Once emotions cooled, several members of his team acknowledged it was the right call. Others never did. Going back over it brought Lesmes to tears. It isn’t exactly the fairy-tale ending to a good decision. Had NASA canceled the launch, Allan McDonald told me that engineers who pushed to abort might have been cast as “Chicken Littles.” Chicken Little doesn’t fare well in the space business. As NASA engineer Mary Shafer once articulated, “Insisting on perfect safety is for people who don’t have the balls to live in the real world.” It is no wonder that organizations struggle to cultivate experts who are both proficient with their tools and prepared to drop them. But there is an organizational strategy that can help. The strategy, strange as it sounds, is to send a mixed message.

“Congruence” is a social science term for cultural “fit” among an institution’s components—values, goals, vision, self-concepts, and leadership styles. Since the 1980s, congruence has been a pillar of organizational theory. An effective culture is both consistent and strong. When all signals point clearly in the same direction, it promotes self-reinforcing consistency, and people like consistency.

Plenty of profiles of individual businesses were written in support of congruence. But in the first study that systematically examined a broad swath of organizations across an industry, researchers who studied cultural congruence at 334 institutions of higher education found that it had no influence on any measure of organizational success whatsoever.

Administrators, department heads, and trustees in strongly congruent institutions did have an easier time classifying the culture when asked, but there was no impact at all on performance, from the academic and career development of students to the satisfaction of faculty and the financial health of the college. The researcher who led that work went on to study thousands of businesses. She found that the most effective leaders and organizations had range; they were, in effect, paradoxical. They could be demanding and nurturing, orderly and entrepreneurial, even hierarchical and individualistic all at once. A level of ambiguity, it seemed, was not harmful. In decision making, it can broaden an organization's toolbox in a way that is uniquely valuable.

Philip Tetlock and Barbara Mellers showed that thinkers who tolerate ambiguity make the best forecasts; one of Tetlock's former graduate students, University of Texas professor Shefali Patil, spearheaded a project with them to show that cultures can build in a form of ambiguity that forces decision makers to use more than one tool, and to become more flexible and learn more readily.

In one experiment, subjects played the role of corporate human resources managers who had to predict the performance of job applicants. The managers were presented with a standard evaluation process that showed them how a candidate's skills were typically weighted, and then told that they would be evaluated (and paid) based on how they made decisions. In a sped-up simulation of real life, after each prediction they could see how the candidate actually performed according to company records. In some batches of applications, the candidates performed as the standard evaluation process predicted; in others, they weren't even close. Yet, over and over, the individual managers conformed to standard procedure no matter what the results told them, even when it clearly was not working, and even when a better system was easily discoverable. They failed to learn with experience. Until a wrinkle was added. Conformist managers were given fake *Harvard Business Review* research proclaiming that successful groups prioritize independence and dissent. Miraculously, their minds were opened and they started learning. They began to see when the standard evaluation process clearly needed to be modified or discarded. They were learning with experience, and their predictions became more accurate. The managers were benefitting from *incongruence*. The formal, conformist company process rules were balanced out by an informal culture of individual autonomy in decision making and dissent from the typical way of doing things.

Incongruence worked in the other direction as well. HR managers who were given a standard evaluation process but told that only the accuracy of their predictions mattered began by ditching the process and making up their own rules. They never learned when the standard process did indeed work. In that case, the cure was fake *Harvard Business Review* research indicating that successful groups prioritize cohesion, loyalty, and finding common ground. Again, the HR managers became learning machines; they suddenly hewed closer to the traditional process when it had value, but continued to deviate readily when it didn't, as NASA should have.

Business school students are widely taught to believe the congruence model, that a good manager can always align every element of work into a culture where all influences are mutually reinforcing—whether toward cohesion or individualism. But cultures can actually be too internally consistent. With incongruence, "you're building in cross-checks," Tetlock told me.

The experiments showed that an effective problem-solving culture was one that balanced standard practice—whatever it happened to be—with forces that pushed in the opposite direction. If managers were used to process conformity, encouraging individualism helped them to employ "ambidextrous thought," and learn what worked in each situation. If they were used to improvising, encouraging a sense of loyalty and cohesion did the job. The trick was expanding the organization's range by identifying the dominant culture and then diversifying it by pushing in the opposite direction.

By the time of the *Challenger* launch, NASA's "can do" culture manifested as extreme process accountability combined with collectivist social norms. Everything was congruent for conformity to the standard procedures. The process was so rigid it spurned evidence that didn't conform to the usual rules, and so sacred that Larry Mulloy felt protected by a signed piece of paper testifying that he had followed the usual process. Dissent was valued at flight readiness reviews, but at the most important moment, the most important engineering group asked for an offline caucus where they found a way, in private, to conform. Like the one engineer said, without data, "the boss's opinion is better than mine."

The more I spoke with Captain Lesmes, the more it seemed to me that he had felt strongly outcome accountable—searching for a solution even if it deviated from standard procedure—within an extraordinarily potent collective culture that ensured he would not make the decision to deviate easily. He had, as Patil, Tetlock, and Mellers wrote, harnessed "the power of cross-pressures in promoting flexible, ambidextrous thought." The subtitle of that paper: "Balancing the Risks of Mindless Conformity and Reckless Deviation."

Superforecasting teams harnessed the same cultural cross-pressure. A team was judged purely by the accuracy of its members' forecasts. But internally the Good Judgment Project incentivized collective culture. Commenting was an expectation; teammates were encouraged to vote for useful comments and recognized for process milestones, like a certain number of lifetime comments.

Prior to *Challenger*, there was a long span when NASA culture harnessed incongruence. Gene Kranz, the flight director when Apollo 11 first landed on the moon, lived by that same mantra, the valorized process —"In God We Trust, All Others Bring Data"—but he also made a habit of seeking out opinions of technicians and engineers at every level of the hierarchy. If he heard the same hunch twice, it didn't take data for him to interrupt the usual process and investigate.

Wernher von Braun, who led the Marshall Space Flight Center's development of the rocket that propelled the moon mission, balanced NASA's rigid process with an informal, individualistic culture that encouraged constant dissent and cross-boundary communication. Von Braun started "Monday Notes": every week engineers submitted a single page of notes on their salient issues. Von Braun handwrote comments in the margins, and then circulated the entire compilation. Everyone saw what other divisions were up to, and how easily problems could be raised. Monday Notes were rigorous, but informal.

On a typed page of notes from two days after the moon landing in 1969, von Braun homed in on a short section in which an engineer guessed why a liquid oxygen tank unexpectedly lost pressure. The issue was already irrelevant for the moon mission, but could come up again in future flights.

“Let’s pin this down as precisely as possible,” von Braun wrote. “We must know whether there’s more behind this, that calls for checks or remedies.” Like Kranz, von Braun went looking for problems, hunches, and bad news. He even rewarded those who exposed problems. After Kranz and von Braun’s time, the “All Others Bring Data” process culture remained, but the informal culture and power of individual hunches shriveled.

In 1974, William Lucas took over the Marshall Space Flight Center. A NASA chief historian wrote that Lucas was a brilliant engineer but “often grew angry when he learned of problems.” Allan McDonald described him to me as a “shoot-the-messenger type guy.” Lucas transformed von Braun’s Monday Notes into a system purely for upward communication. He did not write feedback and the notes did not circulate. At one point they morphed into standardized forms that had to be filled out. Monday Notes became one more rigid formality in a process culture. “Immediately, the quality of the notes fell,” wrote another official NASA historian.

Lucas retired shortly after the *Challenger* disaster, but the entrenched process culture persisted. NASA’s only other fatal shuttle accident, the space shuttle *Columbia* disintegration in 2003, was a cultural carbon copy of the *Challenger*. NASA clung to its usual process tools in an unusual circumstance. The *Columbia* disaster engendered an even stronger ill-fated congruence between process accountability and group-focused norms. Engineers grew concerned about a technical problem they did not fully understand, but they could not make a quantitative case. When they went to the Department of Defense to request high-resolution photographs of a part of the shuttle they thought was damaged, not only did NASA managers block outside assistance, but they apologized to DoD for contact outside “proper channels.” NASA administrators promised the violation of protocol would not happen again. The *Columbia* Accident Investigation Board concluded that NASA’s culture “emphasized chain of command, procedure, following the rules, and going by the book. While rules and procedures were essential for coordination, they had an unintended negative effect.” Once again, “allegiance to hierarchy and procedure” had ended in disaster. Again, lower ranking engineers had concerns they could not quantify; they stayed silent because “the requirement for data was stringent and inhibiting.”

The management and culture aspects of the *Challenger* and *Columbia* disasters were so eerily similar that the investigation board decreed that NASA was not functioning as “a learning organization.” In the absence of cultural cross-pressures, NASA had failed to learn, just like the subjects in Patil’s work who were placed in strongly congruent cultures.

There were, though, individuals in NASA who learned vital culture lessons, and when the time came, put them to use.

• • •

In the spring of 2003, just two months after NASA lost the space shuttle *Columbia*, it had to decide whether to scrap a high-profile project that had been forty years and three-quarters of a billion dollars in the making. Gravity Probe B was a technological marvel designed for a direct test of Einstein’s general theory of relativity. It would be launched into space to measure how Earth’s mass and rotation warped the fabric of space-time, like a bowling ball twirling in a vat of honey. GP-B had the distinction of being the longest-running project in the history of NASA. That was not a compliment.

It was conceived one year after the founding of NASA itself. The launch was delayed numerous times for technical problems, and the project was nearly cancelled on three separate occasions. There were staff members at NASA who no longer thought its mission was possible, and funding had to be rescued repeatedly by a Stanford physicist with a knack for lobbying Congress.

The technological challenges were immense. The probe required the roundest objects ever manufactured—quartz gyroscope rotors the size of ping-pong balls and so perfectly spherical that if you blew them up to the size of Earth, the highest mountain peak would be eight feet tall. The gyroscopes had to be cooled to −450°F by liquid helium, and the probe required surgically delicate thrusters for precise maneuvering. The technology took twenty years in development before it was ready for a test flight.

Congressional eyes were on NASA. The agency could not afford to launch the probe and have a high-profile failure right after *Columbia*. But if the Gravity Probe B launch had to be delayed once more, it could be the last time. “There was a *huge* amount of pressure to get this thing flown,” Rex Geveden, the GP-B program manager, told me. Unfortunately, engineers preparing for the prelaunch flight readiness review found a problem.

The power supply to an electronics box was interfering with a critical scientific instrument. Thankfully, the box only had to work at the beginning of the mission, to get the gyroscopes spinning. It could then be turned off, so it was not a catastrophic issue. But it was unexpected. If there were other flaws that prevented the box from spinning up the gyroscopes to start the experiment, the mission would be a total waste.

The giant Thermos-like container holding the gyroscopes had already been filled with liquid helium, cooled, and sealed for launch. If the box needed inspection, parts that had taken three months to install would have to come off the probe; a launch delay would cost $10–$20 million. Some engineers felt there was more risk in removing and potentially damaging parts than in leaving it all alone. Stanford University was the prime contractor, and the Stanford team leader “was confident that we could succeed,” he said, “so I pushed hard that we should go ahead and fly.” NASA’s chief engineer and head scientist for Gravity Probe B also both pushed to launch. Plus, the probe had been moved to Vandenberg Air Force Base in California for launch, and a delay would increase the chance of GP-B sitting there when an earthquake struck. So: race, or don’t race?

The decision was in Geveden’s hands. “My God, I can’t even express how stressful it was,” he told me. Even before the latest snafu, he had a hunch held lightly—he was uneasy about how the electronics box had been managed. But as long as the box was attached to the probe, there would be no more information forthcoming.

Geveden joined NASA in 1990, and was a keen observer of the culture. “When I was coming through NASA,” he said, “I had the intuition that there’s a real conformance culture.” Early in his tenure, he attended a team-building class offered by the agency. On the very first day the instructor asked the class, rhetorically, for the single most important principle in decision making. His answer: to get consensus. “And I said, ‘I don’t think the people who launched the space shuttle *Challenger* agree with that point,’” Geveden told me. “Consensus is nice to have, but we shouldn’t be optimizing happiness, we should be optimizing our decisions. I just had a feeling all along that there was something wrong with the culture. We didn't have a healthy tension in the system." NASA still had its hallowed process, and Geveden saw everywhere a collective culture that nudged conflict into darkened corners. "You almost couldn't go into a meeting without someone saying, 'Let's take that offline,'" he recalled, just as Morton Thiokol had done for the infamous offline caucus.

Geveden, in his own way, was in favor of balancing the typical, formal process culture with a dose of informal individualism, as Kranz and von Braun once had. "The chain of communication has to be informal," he told me, "completely different from the chain of command." He wanted a culture where everyone had the responsibility to protest if something didn't feel right. He decided to go prospecting for doubts.

He deeply respected Stanford's electronics manager. The manager had worked with the same kind of power supply before, and viewed it as fragile technology. After a formal meeting in which NASA's head engineer and its head scientist on the project both advocated for leaving the box in place, Geveden held informal individual meetings. In one of those, he learned from a member of the NASA team that a manager from Lockheed Martin, which had built the box, was concerned. Like *Challenger*'s O-rings, the known problem with the box was surmountable, but it was unexpected. There were unknown unknowns.

Against the recommendation of the chief engineer and the Stanford team leader, Geveden decided to scrub the launch and pull the box. Once it came off, engineers quickly discovered three other design problems that had not been clear in schematics, including a case of having used the flat-out wrong parts. The surprises prompted Lockheed to go back over every single circuit in the box. They found twenty separate issues.

As if Gravity Probe B was required by the space gods to scale every imaginable obstacle, a month after the box was pulled there was an earthquake near the launch site. The launch vehicle was slightly damaged, but fortunately the probe was intact. Four months later, in April 2004, GP-B finally took off. It was the first direct test to support Einstein's idea that Earth drags the fabric of space-time around with it as it spins. The technology left a greater legacy. Components designed for Gravity Probe B improved digital cameras and satellites; the centimeter-accurate GPS was applied to automatic aircraft landing systems and precision farming.

The following year, a new NASA administrator was appointed by the president. The new administrator demanded the kind of individualism and opinionated debate that could serve as a cross-pressure for NASA's robust process accountability. He made Geveden the associate administrator, essentially the COO of NASA, and the highest position in the agency that is not politically appointed.

In 2017, Geveden took his lessons to a new role as CEO of BWX Technologies, a company whose wide purview includes nuclear propulsion technology that could power a manned Mars mission. Some of BWX Technologies' decision makers are retired military leaders whose dearly held tool is firm hierarchy. So when Geveden became CEO, he wrote a short memo on his expectations for teamwork. "I told them I expect disagreement with my decisions at the time we're trying to make decisions, and that's a sign of organizational health," he told me. "After the decisions are made, we want compliance and support, but we have permission to fight a little bit about those things in a professional way." He emphasized that there is a difference between the chain of command and the chain of communication, and that the difference represents a healthy cross-pressure. "I warned them, I'm going to communicate with all levels of the organization down to the shop floor, and you can't feel suspicious or paranoid about that," he said. "I told them I will not intercept your decisions that belong in your chain of command, but I will give and receive information anywhere in the organization, at any time. I just can't get enough understanding of the organization from listening to the voices at the top."

His description reminded me of Girl Scouts CEO Frances Hesselbein's "circular management." Instead of a ladder, the organizational structure was concentric circles, with Hesselbein in the middle. Information could flow in many directions, and anyone in one circle had numerous entry points to communicate with the next circle, rather than just a single superior who acted as a gate. When she explained it to me, it seemed a lot like the kind of incongruence Geveden worked to engender, and the kind that Captain Lesmes wielded: a differentiated chain of command and chain of communication that produced incongruence, and thus a healthy tension. An occasionally confusing but effective mix of strong formal and informal culture. A trio of psychology and management professors who analyzed a century of Himalayan mountain climbers—5,104 expedition groups in all—found that teams from countries that strongly valued hierarchical culture got more climbers to the summit, but also had more climbers die along the way. The trend did not hold for solo climbers, only teams, and the researchers argued that hierarchical teams benefitted from a clear chain of command, but suffered from a one-way chain of communication that obscured problems. The teams needed elements of both hierarchy and individualism to both excel and survive.

It is a difficult balancing act, cultivating aspects of a culture that seem on their face to push against one another. There are no rules for the qualitative hunches of space shuttle engineers or pararescue jumpers lacking intel. Incongruence, as the experimental research testified, helps people to *discover* useful cues, and to drop the traditional tools when it makes sense.

Karl Weick's tools insight reminded me of an experience I had as a graduate student, working aboard the Research Vessel *Maurice Ewing* in the Pacific Ocean. The ship was bouncing sound waves off the ocean floor to image underwater volcanoes. I got to know a few volcano experts who truly saw the world through volcano-colored glasses. Despite ample evidence that an asteroid impact was either the primary cause of the dinosaur extinction, or at least very important, they insisted that volcanic eruptions were clearly the real culprit. If anything, one told me, the asteroid was really just a lucky knockout punch; volcanoes had already delivered the body blows. He seemed to attribute a whole slew of mass extinctions to volcanoes, some with compelling evidence, others with pretty much none. When all you have is a volcanologist, I learned, every extinction looks like a volcano. That is not necessarily bad for the world. They *should* challenge accepted wisdom, and it drives those narrowly focused experts to find volcano knowledge where no one else is looking. But when entire specialties grow up around devotion to a particular tool, the result can be disastrous myopia.

Interventional cardiologists, for example, specialize in treating chest pain by placing stents—a metal tube that pries open blood vessels. It makes a ton of sense: a patient comes in with chest pain, imaging shows a narrowed artery, a stent is placed to open it and preclude a heart attack. The logic is so compelling that a prominent cardiologist coined the term "oculostenotic reflex," from the Latin for "eye," and stenotic, from the Greek for "narrow," meaning: if you see a blockage, you'll reflexively fix a blockage. Except, repeatedly, randomized clinical trials that compared stents with more conservative forms of treatment show that stents for patients with stable chest pain prevent zero heart attacks and extend the lives of patients a grand total of not at all.

The interventional cardiologists are seeing and treating one tiny part of a complicated system; the cardiovascular system isn't a kitchen sink, and it turns out that treating one blocked pipe often doesn't help. Plus, about one in fifty patients who get a stent will suffer a serious complication or die as a result of the implantation procedure. Despite the bird's-eye evidence, cardiologists who specialize in using that tool reported that they simply cannot believe that stenting doesn't work, even when their compensation was not tied to performing the procedure. Being told to stop using stents was like being told to forget you are an interventional cardiologist. The instinct, often well-meaning, to use interventions that seem logical but that have not been shown to help may explain the finding of a 2015 study: patients with heart failure or cardiac arrest were less likely to die if they were admitted during a national cardiology conference, when thousands of top cardiologists were away. "At large cardiology conventions, my colleagues and I have often joked that the convention center would be the safest place in the world to have a heart attack," cardiologist Rita F. Redberg wrote. "[The conference study] turned that analysis around."

Similarly harrowing findings are now appearing all over medicine, wherever specialties have arisen for the use of a particular tool. One of the most common orthopedic surgeries in the world involves shaving a torn meniscus—a piece of cartilage in the knee—back to its original crescent shape. A patient reports knee pain; an MRI shows a torn meniscus; naturally, a surgeon wants to fix it. When five orthopedic clinics in Finland compared the surgery with "sham surgery"—that is, surgeons took patients with knee pain and a torn meniscus to operating rooms, made incisions, faked surgeries, and sewed them back up and sent them to physical therapy—they found that sham surgery worked just as well. Most people with a torn meniscus, it turns out, don’t have any symptoms at all and will never even know. And for those who do have a torn meniscus and knee pain, the tear may have nothing to do with the pain.

Seeing small pieces of a larger jigsaw puzzle in isolation, no matter how hi-def the picture, is insufficient to grapple with humanity’s greatest challenges. We have long known the laws of thermodynamics, but struggle to predict the spread of a forest fire. We know how cells work, but can’t predict the poetry that will be written by a human made up of them. The frog’s-eye view of individual parts is not enough. A healthy ecosystem needs biodiversity.

Even now, even in endeavors that engender specialization unprecedented in history, there are beacons of breadth. Individuals who live by historian Arnold Toynbee’s words that “no tool is omnicompetent. There is no such thing as a master-key that will unlock *all* doors.” Rather than wielding a single tool, they have managed to collect and protect an entire toolshed, and they show the power of range in a hyperspecialized world.