Performance Hacking: The Contagious Business Practice that Corrodes Corporate Culture, Undermines Core Values, and Damages Great Companies

Robert D. Austin
Richard L. Nolan

Performance Hacking: The Contagious Business Practice that Corrodes Corporate Culture, Undermines Core Values, and Damages Great Companies

Robert D. Austin
Ivey Business School

Richard L. Nolan
Harvard Business School

Working Paper 21-003

Introduction

August 7, 1955 is an important date in commercial aviation history. You could say it began the jet airliner age, though other dates might also qualify. Jet engines had proven successful in military aircraft. But no one knew then whether members of the general public would ever want to fly in airplanes without propellers.

The first jet airliner, the Dehavilland DH-106 Comet, had gotten things off to a bad start. In its first twelve months of service, spanning late 1952 and early 1953, the Comet failed three times. On two of these occasions, everyone onboard – 54 people in all – perished. The first two incidents were attributed to pilot error, but evidence in the third case suggested structural failure in mid-flight. Two more crashes, both mid-air breakups, followed in 1954, killing 56 more people, and the Comet was withdrawn from service.[1] These spectacular and highly publicized failures did little to inspire confidence in this new mode of transport.

In 1955, against this discouraging backdrop, Bill Allen, CEO of the Boeing Company, put into motion a plan to turn popular opinion of jet travel to a more positive bearing. He didn't have a choice, really. Way back in 1952, before Dehavilland's problems with the Comet, Allen had placed a bet. He'd convinced Boeing's board of directors that jets were the future and to invest in the development of a bold new jet airliner.[2]

It was a huge bet. The money required equaled all the profits earned by the company since the end of World War II. And it was a risky bet. No one had asked for this new plane. Boeing developed most projects in those days in response to government requests, or after a clear expression of willingness to buy from the airlines. Allen's new project enjoyed neither of these advantages. Demand for the plane at levels that would pay back its vast development cost remained hypothetical.[3] In effect, Allen had bet the company on his vision of the future.

In this way, on that sunny August day in 1955, the prospects for his vision came to head, with the debut of the Boeing 367-80 – the "Dash 80" as it was called within the company. The official program called for a demonstration flight. Boeing's Chief Test Pilot Alvin "Tex" Johnston, accompanied by co-pilot Jim Gannett and test engineer Bel Whitehead, would fly the Dash 80 over the 1955 Gold Cup hydroplane races at Seattle's Lake Washington. The 250,000 or so spectators in attendance that Sunday included airline executives from around the world, in Seattle for a meeting of the International Air Transport Association. How the crowds and executives reacted to what they saw would largely determine whether Allen's big, risky bet might pay off – whether Boeing had a fighting chance to overcome the dismal legacy of the Dehavilland Comet. It was make or break time for the Dash 80 project – and, in all likelihood, the Boeing Company.

Allen knew the stakes were high. But there was something he didn't know that day, something that would have sent his stress levels soaring to even greater heights had he known it. Another wildcard was about to enter into the mix. Test pilot Johnston had a secret plan of his own, one that Allen would never have approved.

Johnston's secret, which he revealed to his co-pilot and test engineer only after the plane lifted off the tarmac at Boeing Field, was this: he intended not just to fly the plane over the crowd, but also to execute an impressive, dangerous-looking stunt, a barrel roll. At a speed of almost 500 miles per hour, he would fly the 124-ton experimental aircraft *upside down*. He proceeded to roll the plane not once, but twice. People watching from below oohed and aahed.[4] Whitehead, Johston's flight engineer, took a now famous photo showing the plane's engine sticking uncannily, weirdly upward, upside down over Lake Washington. Later, Johnston explained his confidence in the plan: "It's a one G-force maneuver, that is not at all hazardous, but very impressive." The

forces on the plane, he said, were exactly the same as in level flight : "the airplane never knows it's inverted."[5]

But the unexpected maneuver took Allen by surprise. He recoiled in horror, thinking that the pilot had lost his mind or the airplane was having major difficulties. Standing amid a cluster of airline executives, Allen turned to Larry Bell of Bell Aircraft, who took medicine for a heart condition, and said "Give me one of those damned pills. I need it worse than you do."[6]

Allen was furious. In his reckoning, the purpose of the flyover, to show that this jet airliner was stable and safe, had been compromised by Johnston's grandstanding. After the Comet mishaps, the world needed to see jet airliners as unexciting, reliable transportation, he thought, not as props for use in barnstorming stunts. He summoned Johnston to his office the very next morning.

"What the hell did you think you were doing?" he asked.

"Selling airplanes!" was Johnston's enthusiastic response. He went on to explain how the maneuver entailed zero risk.

"You know that, and now we know that," Allen replied, "But don't do it again."

Despite this expression of displeasure, Allen likely sensed already that morning that Johnston's improvisation might well pay off. At a dinner Allen hosted at his home the evening before, Eddie Rickenbacker, the famed World War I flying ace, who had become the CEO for Eastern Airlines, complained to Allen that he was not invited to ride along on the barrel roll flight. Laughing and shoving Johnston's trademark Stetson hat down over his ears, Rickenbacker called Johnston a "slow rolling SOB."[7] Excitement spread, through the dinner crowd and the industry. Within a month, orders for the commercial version of the plane, the Boeing 707, began rolling in. Boeing would sell more than a thousand 707s before it ceased production of the plane in 1994.

Performance – Real and Professed

We recount the story of Tex Johnston's famous barrel roll to make a point about performance and representations about performance. The entire aim of the high-stakes exercise on that August day in 1955 was to present a favorable impression of the performance of a type of airplane that people had, up until then, been very uncertain about. Johnston and Allen both wanted the audience to see the plane's performance in a favorable light, and planned with that in mind. Johnston enacted a stunt that had very little to do with how the plane would ever be flown in actual service – literally "spinning" to create confidence in the flight-worthiness of the new plane. Neither Allen nor Johnston were even attempting to be unbiased. They staged the entire event to promote an interpretation of the plane's performance that they favored.

But they also weren't faking anything. If the plane had not been capable of the aspects of performance they were presenting that day, the Dash 80 might have gone the way of the Dehavilland Comet, and Johnston and his airborne colleagues might not have lived to tell about it.

The point we want to make here is that test piloting is perhaps the best metaphor for how business leaders ought to behave when it comes to the performance of their own organizations. It's completely responsible for business executives to promote, profess about, and even stage events to demonstrate the capabilities of their organization, the worthiness of its products and services, and the rightness of their managerial actions. Indeed, it would be irresponsible not to.

But there is a question of *balance*. Executive promoting, professing, and staging crosses a line when it represents performance that has little or no basis in empirical reality. Tex Johnston, his flamboyant appearance and reputation notwithstanding, could not extend enthusiastic claims and demonstrations of the 707's performance beyond what was empirically possible without endangering the plane and himself. The laws of physics constrained him, as they do all test pilots.

This didn't prevent him from staging a demonstration that suggested great performance, but it did keep those claims in line with reality.

In their day-to-day duties, business executives are not so reality constrained. They have only their own consciences and the influence of their boards, peer executives, and employees to keep in check any tendencies they might have toward overclaiming. But as we will show, tendencies to misrepresent performance, or rely on hypotheses about performance that are not grounded in empirical reality, can crash a company as readily as it can crash a plane.

Balance as an Executive Virtue

In the *Nicomachean Ethics*, Aristotle constructs a philosophy of individual virtue that relies on the idea of balance. The main idea is that many elements of human character and behavior, including those the we view favorably, can be considered virtuous only when maintained in a state of balance. So, for example, courage is generally considered a positive human character trait, but only up to a point. If you possess too little courage, you are, of course, cowardly. The failing of virtue in that case is obvious. But, as Aristotle argues, virtue can also fail in an opposite manner. *Too much* courage makes you foolhardy, prone to take stupid risks. Virtue lies in keeping an appropriate balance.

We suggest that an executive's stance in representing performance must be guided by a similar sense of balance. When executives are too little oriented toward professions of success, they fail to claim credit where credit is due, which equates to a certain kind of incompetence or irresponsibility. On the other hand, too much or over zealous advocacy of positive interpretations, to the point of detachment from actuals, tip over into a negative range – into what we call *performance hacking*, or *p-hacking*, for short.

The idea of business p-hacking owes an explicit debt to the work of Joseph Simmons, Leif Nelson, and Uri Simonsohn on "p-hacking" in research. P-hacking in its original meaning refers to the practice of reanalyzing and reshaping data, or gathering additional data, until a study yields a statistically significant result (usually indicated by a "p-value" of less than 0.05). It's a form of cherry picking that allows researchers to reach a conclusion they were hoping to reach. Statistically significant results are the path to publication and promotion in academic research, so incentives to p-hack are strong. But such "data dredging" practices are anathema to the appropriate conduct of science, because results reached by p-hacking are not replicable. They can't be reproduced by other scientists because they don't report real effects.

Except in the most egregious cases of p-hacking in research, the problem is not obvious when it's occurring. The methodologies that guide how research is conducted inevitably involve many small decisions. And while these need to be defensible to the peer reviewers, who decide what gets published and doesn't, there are gray zones and matters of judgment. Simmons, Nelson, and Simonsohn have shown is that there are telltale signs in aggregate, patterns in the parameters of ongoing analyses, that suggest that p-hacking might be occurring, even in research that has passed muster in peer review and been published. P-hacking, then, is not easy to detect. So, ultimately, whether research proceeds in a manner that assures scholarly integrity and replicable results depends, to a large extent, on the researcher's judgment and conscience. Good researchers strive to maintain a virtuous balance as they act on their professional curiousity. In each small decision, they guard against the possibility that eagerness to discover new truths doesn't tip over into bias and manipulation.

P-hacking in business is similar. There is, of course, no objective standard of performance in business. There are numerous different measures – stock price, sales and profit numbers, return

on sales, return on equity, and many, many others. But as in research, arriving at any these requires a multitude of small decisions, how to account for various factors and at what time, judgments about individual components and inputs in calculation. And there are many gray areas. In business, the measures of performance put forward do get scrutinized, not by peer reviewers, but by analysts and others seeking to make wise investments in specific companies. In business, as in research, there are strong incentives to p-hack; executive bonuses often depend on specific measures of performance. Observers are sometimes able to point out patterns suggestive of business p-hacking, such as, for example, the use of questionable means to smooth numbers and control their trends over time as "earnings management," which have been extensively researched.[8] There is a lot of research on earnings management; but much of it focuses on instances serious enough to attract enforcement actions from regulators.

Our concern here is the vast majority of cases of p-hacking in business that are more subtle and less obvious, harder (if not impossible) to pin down objectively at the moment of their occurrence. We submit, however, that executives who engage in p-hacking do, or should, realize that they are doing it. Some examples become clear in retrospect, and in those instances it often true that executive knew what they were doing. And p-hacking is, in the final analysis, no less consequential in business than in research. When p-hacking goes on for long enough, there is a bill that will need to be paid.

Contrast that day in 1955, when Boeing launched the jet airline age, with a day at the same company almost 52 years later, a debut event for a different airplane. The 787 Dreamliner was being presented to the world for the first time, rolled out of its hanger on live TV, broadcast on the Today Show hosted by news anchor Tom Brokaw.[9] The date was July 8, 2007 (7/8/7), and the first flight of the new plane was scheduled for less than a year later, in the second quarter of 2008. But

unlike the 1955 demonstration of the Dash 80, this 2007 event was all show and no real performance. Unknown to the world at large, but all too well known to Boeing workers, the announced first flight date was pure fantasy. The prototype rolled out for TV viewers was a stage prop, as rumor had it, "cabled together with baling wire." Major components were in widely varying condition, many far short of required specifications.[10] The 787 program was in serious trouble and Boeing executives knew it. But that was not the message company executives were professing that day.

Already, in 2007, the contagion we call p-hacking, had infected Boeing. It entered the organization in the behavior of senior leaders. It spread, passing gradually through what we call "the five stages of p-hacking," corroding a unique corporate culture that had held together for decades, chipping away at northstar values that had made this great company successful for more than 100 years – most notably, the sacred axiom traceable all the way back to Bill Boeing himself: "Build to perfection."

The full extent of the damage would not become obvious for another 12 years. In 2018 and 2019, a different new airplane, the 737 MAX, would crash twice within six months. A total of 346 people would lose their lives on the two flights, Lion Air flight 610 and Ethiopian Airlines flight 302. Subsequent investigation would reveal that problems that caused the crash were known within Boeing, and that opportunities to address them were suppressed as part of elaborate efforts to maintain strained financial performance representations. The cynicism evident in Boeing employee emails released in early January 2020 showed conclusively that p-hacking at Boeing had reached its final stage, in which an artificial performance reality completely supplants more realistic assessments, as a context and guide for executive action.

We believe that executives can and should be self-aware enough to know when they are engaging in p-hacking. But to be fair, we acknowledge that p-hacking has become largely normalized in much of today's business. The means that enable p-hacking are taught in business schools – a fact that implicates us, the two business school professors authoring this book. In fact, the rise of p-hacking has been accompanied by enabling ideas that have infused business school curricula, for example, the idea that a good manager can manage anything – or, more specifically, that "managing by the numbers" creates comparability that allows managers to transfer their expertise across very different businesses and industries. Such comparability is, we suggest, an illusion, and far from a harmless one. For it is the analytical machinery designed to summarize performance abstractly, aimed at permitting this kind of comparability, that provides many degrees of freedom for p-hacking. Performance management innovators, in companies and in academia, aimed for comparability across businesses and industries but missed; the abstractions they developed in effort to achieve comparability became the enablers of the false performance realities that resulted from p-hacking.

Another enabling idea is that it is right and good to design incentives based on abstract performance metrics. The p-hacking this idea enables now routinely allows executives to enrich themselves to a degree unprecedented historically. When you combine analytical machinery that allows manipuation with strong incentives to manipulate, you should perhaps not be surprised if a business crashes from time to time

We concede that it is a part of an executive's job to inspire employees. That this often requires being optimistic, sometimes to the point of incredulity. And that this behavior can have positive effect. Steve Jobs was notorious for his "reality distortion field." "Fake it 'til you make it," has become a way of urging persistence. Marketers and communications specialists have a

responsibility to manage a company's image and brand, and doing that well is an important part of how companies maintain customer loyalty and healthy profit margins. This is why p-hacking is difficult to detect and challenging to eradicate. It begins with and mixes with good ideas and good intentions, with optimism, not fakery. But somewhere along the way, if executives are not mindful, a line can be crossed. The balance that keeps executive behavior virtuous in the Aristotelian sense can be lost.

Achieving virtue is not supposed to be easy. But when we look to the leaders of economic entities of historically unprecedented size and wealth, we are right to expect virtue. Framing this issue as a matter of virtue is, we believe, appropriate. This is our effort to explain the consequences of such failures of leadership virtue, and also how such failures might be prevented or undone.

The Five Stages of P-Hacking

P-hacking is especially damaging when its effects diffuse within an organization's culture, undermining core values and destroying organizational capability. When leaders hack performance with enough frequency and vehemence, employees get it. They understand that what they are expected to do in their jobs has changed. They see that, going forward, their bosses will care less about getting the job done according to past definitions, and that other, more abstract representations of performance are ascendant. Within the organization's ranks, substantive innovators lose out to innovators of appearances, those willing to exploit gray areas to bend assessments and tailor reports. Simplicity and transparency give way to complexity, and a corrosive cynicism takes hold. Table 1 describes the progression of this affliction through stages.

The eventual result is loss of organizationl capabilities. This loss can take many forms: the of inability to assure quality, or to create new ways of producing real value or engaging in meaningful organizational change.

Stage	Description	Explanation	Basis of Performance Assessment
1	Orientation toward Actual Performance	Executives are focused on the core business and core values – on driving performance with true measures and decision making from corporate values. Like test pilots, they are constrained to realistic representations of performance (they constrain themselves).	Keeping It Real
2	Abstraction of Performance Criteria	Executives develop more sophisticated ways of measuring performance that examine specific facets of importance; this is largely insight generating and constructive at first; it helps decompose understanding of actual performance; but abstracted sub-component measures of performance do not fully measure actual performance – this fact creates the potential for future p-hacking	Analyzing What's Real
3	Investment of Significant Resources in Abstract Performance Management	Executives drive emphasis on alternative, abstract performance measures throughout the organization; line managers take up and use alternative measurements, often are incentivized by them, and use them in communicating internally and externally; still, these practices are not yet necessarily harmful, as long as core values keep the primary emphasis on actual performance; alternative measurements can, at this stage, be used to point out important ways in which company performance is differentiated from the performance of rivals; this is, up to a point, entirely responsible image management	Losing Track of What's Real
4	Diversion of Major Resources from Actual to Managed Performance Representations	This is where p-hacking *per se* begins; executives realize that alternative, component measures can distract from shortfalls in actual "real" performance. Marshalling sophisticated tools, they increasingly lobby for performance evaluation and compensation based on more abstract and manipulatable alternative measures	Distracting from What's Real
5	P-Hacking Infuses into Corporate Culture	The organizational culture adapts to reflect the new emphasis on abstracted measures, and the de-emphasis of actual performance. Line managers and employees understand that efforts directed to actual performance – e.g., building great airplanes or industrial equipment, or inventing great new technologies – at this point abstract performance begins to undermine real performance, and abstract performance measures can distort and it becomes harder to align rewards for actions that align reflect historical core values	Artificial Reality

Table 1: The Five Stages of P-Hacking

We unfold our explanation of what we call p-hacking through the stories of three iconic, 100-year old companies: Boeing, General Electric (GE), and IBM. These companies have seen technologies come and go, and the world rocked by many major events over long periods of time. Despite all this, they have mostly been able to thrive. When Jeffery Immelt, then the Chief Executive Officer of GE, addressed the Detroit Economic Club in 2009, he explained his company's long-term success in a way that could be applied to any of these three companies:

> We are the oldest remaining company in the Dow Jones industrial average. This not because we are perfect company; it is because we adapt. Through the years, we have remained productive and competitive.[1]

Management professor Vicki TenHaken, author of *Lessons from Century Club Companies*, contends that long-surviving firms "have very strong cultures and are deliberate about preserving them. Often the beliefs and values of the founder have been passed on through generations."[2] As we will see, this has largely been true of Boeing, GE, and IBM. All three have approached 100-year anniversaries in recent years with many reasons to be proud.

And yet recently – in the last 20 years or so – it appears that something has changed for each of them. Something about recent efforts by these firms to adapt has been different. Despite their long and successful pasts, each of these great firms has faltered, and to no small degree.

---

[1] Text of Immelt's speech, *Wall Street Journal*, June 26, 2009, https://www.wsj.com/articles/SB124603518881261729, accessed on October 27, 2019.
[2] Vicki TenHacken, "How Company's Survive 100 Years, *Crain's Detroit Business*, September 4, 2016, https://www.crainsdetroit.com/article/20160904/BLOG200/160909939/how-companies-survive-for-100-years, accessed on October 27, 2019.

Boeing - Grounded

Throughout its history, Boeing has dealt with many crises. Some were externally inflicted: anti-trust actions by the US government in the 1930s, the economic conditions in the aftermath of world war, and the oil crisis of the early 1970s. Others followed from challenging ideals set by founder Bill Boeing, which called for the company to "let no airplane technology pass us by" and "build to perfection." Boeing president Clairmont Egtvedt cemented these principles into the so-called "big Boeing" pattern in the 1930s. The pattern involved regularly betting the company on ambitious new airplane projects that incorporated big technological advances. The Boeing 247, B-17, Dash-80, 707, 737, 747 were all game-changing airplanes. But the costs of incorporating new technologies into them and building them to perfection strained the company every time, sometimes to near the breaking point. And yet Boeing has repeatedly survived its big bets to rake in big returns from often having the best products in the industry. Thoughout much of the life of the company, Boeing has dominated its competitive landscape.

Boeing's recent stumbles, however, centered on its two newest airplanes, the 787 Dreamliner and the 737 MAX, seem different from the company's past struggles. The 787 program was incredibly ambitious, both in terms of technology in the plane and how Boeing intended to build it. The 787 project, as originally conceived, was another bet-the-company initiative, a grand effort to transform the company and its ways of doing business to arrive at an unprecedented technological accomplishment. At first, the 787 project fit the Big Boeing pattern. If the project had been carried out as conceived in the late 90s and early 00s by then-CEO Phil Condit, it is likely that it would have again suffered the cost overruns and other challenges involved in such a big bet. They might have again veered close to the brink of corporate viability. But it is also likely, we suggest, that the difficulties that have beset the company since then might have been avoided.

According to many, including many insiders, Boeing's acquisition of McDonnell Douglas in the late 90s significantly changed the culture of the company, from the Bill Boeing inspired engineering orientation, to a *financial* engineering orientation. Spectacular experiments in product development, engineering, information technology, and operational systems design became, within the new "financialized" executive culture, less about actual technical collaboration and more about risk avoidance. In directing their gaze to favored performance abstractions, the company's leaders took their eyes off the engineering and operational "ball."

Long and expensive delays on the 787 program were one result. More recently, when the evolving financial logic of the new Boeing ruled out development of a new plane as too financially risky, breaking the Big Boeing pattern, they decided to instead modify the existing 737 platform, yielding the 737 MAX. The MAX had much bigger engines, which were not well matched to the legacy 737 airframe design; the larger engines, the way they had to be mounted to the plane, destabilized it in flight, necessitating the now infamous MCAS system software correction, implicated in both fatal crashes.

P-hacking logic kept concerns about the MCAS system quiet, resulting in two avoidable crashes and grounding of the plane. In US Congressional hearings on October 29, 2019, one senator called the planes "flying coffins." At the time of this writing, it remains unclear when the 737 MAX will return to commercial service, and whether the impact of this debacle might yet devastate the company.

Even if the company survives, it is unclear how it will be managed going forward. Within the p-hacking logic of today's Boeing executives, it is probably not possible to take the risks associated with developing a completely new airplane. The managerial culture has evolved to a point where a financial model is more central to the company's activities than building airplanes.

This shows us something else about the practice of p-hacking: its frequent aim is to avoid the appearance of risk and uncertainty, to instead convey performance and managerial competence in terms of relentless, unnatural evenness and consistency.

There is nothing particularly even or consistent in the Big Boeing pattern of betting the company on each new airplane. To translate the extreme cycles of steep development costs early in a program that are recouped only years later, when sales revenues start to come in, requiresnsome formidable financial acrobatics.

One example of such acrobatics, "program accounting," has long been used by Boeing to allow the company to expense estimated average costs over the life of a new airplane project, instead of the actual production costs of an aircraft. Program accounting lowers the effect on profitability of the initially high costs of tooling and manufacturing a new model of aircraft, in essence, smoothing Boeing's profits. But it is based on estimates of sales and costs far into the future, and depends greatly on gray area assumptions, which meant it was, in the words of one analyst, "almost certainly destined to be wrong."[11] Such profit smoothing accomplished much more for a executives striving to project an image of evenness and consistency than it did for outsiders trying to assess the performance of the company, to make investment decisions. Without program accounting, between 2012 and 2016, Boeing would have realized only $1.4 billion in profits for its commercial aircraft division, rather than the $25.2 billion it actually reported.[12] Program accounting has made comparing Boeing's performance to others in the same industry harder, not easier. No one else in the industry currently uses it.

Program accounting at Boeing is not new. And though it represented a step in the direction of a higher stage of p-hacking, it also co-existed with the "build to perfections" Boeing culture instilled by the company's founders. What this shows us is that a company, and its leaders and

culture, often often respond to a mix of influences. While the Bill Boeing culture remained strong, p-hacking remained reasonably well in check, despite occasional movements in that direction, such a program accounting. The temptations of p-hacking are ever present, but movement to the fifth stage of p-hacking is not automatic or inevitable. Executive leadership can use performance abstractions that enable p-hacking, and even benefit the company by their use, without allowing a firm to progress to the later stages of p-hacking. GE, another of the iconic companies on which we focus in this book, provides vivid evidence of this fact.

GE – Ground Zero for P-Hacking, Source of the Boeing Contagion

For decades, GE enjoyed a reputation as one of the world's best managed companies. The company's distinguished history, which traces back to Thomas Edison's invention of the light bulb, is dense with managerial innovations. Before the turn of the 20th century, GE president Charles Coffin had begun a practice for which GE was to become famous, investing heavily in development of future leaders; he also pioneered improvements for hourly workers, such as profit sharing, pensions, and insurance programs, and established the first industrial research laboratory.[13] Owen D. Young, Coffin's successor, was among the first to deploy what would later be called "stakeholder management," described by historian John Winthrop Hammond as "a conception of management, not as an agent of its owners, but as a trustee of all groups vitally interested in industry – owners, employees, and the general public, including customers."[14] In 1956, under CEO Ralph Cordiner, GE established a corporate education center in Crotonville, New York with the explicit purpose of making GE the best managed company in the world.[15] Jack Welch, CEO from 1981 to 2001, became one of the most widely admired business leaders in the world, and was proclaimed "Manager of the (20th) Century" by *Forbes* in 1999. He introduced a

practice of requiring all business units to be number one or two in their respective industries, or else the company would divest itself of that business, and also a widespread roll out of "Six Sigma," a technique for promoting quality and efficiency which yielded major financial benefits for the company.

Not surprisingly, when many companies looked outside for new executives, they often turned to GE. Former GE managers have taken senior roles at many other companies leading to a diaspora of GE know-how. When GE execs have taken leadership roles in other companies, they frequently apply managerial techniques and approaches developed and refined within GE. A significant number of the senior leaders put in charge of Boeing after the McDonnell Douglas merger in 2001 were from GE. Within Boeing, these new leaders followed the pattern, adopting philosophies and practices derived from their GE experiences. And, although applying GE management ideas elsewhere has often led to outstanding leadership and results, we suggest that the potential for p-hacking also travels within some sophisticated GE management approaches.

Some of the impetus behind GE's managerial innovation has come from the company's unusual diversity of business lines within a large conglomerate. Whereas many companies (Boeing and IBM would be good examples) have focused on one industry, GE has, since at least the Ralph Cordiner era, contained a wide variety of business units that compete in very different industries (e.g., power generation equipment, jet engines, medical devices, financial services). Cordiner put in place a decentralized arrangement in which business units were allowed significant autonomy, while senior leaders at headquarters monitored performance and engaged in strategic planning.

The need to assess and compare performance across such diverse business units led to the development of sophisticated abstract tools for analyzing and representing managerial performance. Because these tools needed to work for all business units, they were based on the

premise that performance in management could be measured without reference to a specific business context. That is, measuring business performance was a generic matter, not a matter specific to an industry. A corollary idea, which was gaining currency in business schools around the same time (in the late 1950s and into the 60s) suggested that a "good manager could manage anything," any kind of business. Cross business performance comparability was supposedly enabled be the development of sophisticated quantitive measures. Robert McNamara, first a professor at Harvard Business School, later a senior executive at Ford Motor Company, and then a U.S. Secretary of Defense famously pioneered this way of using numbers to induce comparability. One of the authors, Dick Nolan, worked in the 1960s Defense Department, as one of McNamara's "Whiz Kids," bringing these sophisticated quantitative techniques to the highest levels of the US government.

The new performance abstractions were not immediately harmful. Often they were an excellent source of insights, an impetus to real business improvements. McNammara's efforts within Ford modernized the company, to excellent results. In GE, they were integral to the ways of instruction at Crotonville, and created managers who fanned out through company and brought it notable success. The need for cross-business unit performance comparability became even greater during the Jack Welch era, because of his insistence on ranking and categorizing managers. "A" employees had proven their superior abilities; they were to be rewarded and kept. "B" employees had shown potential; they were to be kept, watched, and nurtured. "C" employees had performed unsatisfactorily; they were to be fired.

Why the process of abstracting performance for comparability creates the potential for p-hacking is pretty intuitive. Abstract representation of performance are many and varied, but do not necessarily summarize the totality of performance. Slicing up performance into facets is productive

of insights, but also raises issues of how to re-integrate the facets into more comprehensive notions of performance. More importantly, all these points of re-integration represent points of opportunity at which the representations of performance can be influenced. Depending on the behaviors that senior leaders model to the organization, these points of influence can be accessed in ways that keep p-hacking in check, or that fail to maintain virtuous balance.

The leadership of Jack Welch himself provides an interesting case. Although he arguably did more to enhance the possibilities of p-hacking than any of his predecessors, by demanding systems of cross business performance comparability, GE does not seem to have moved to the later stages of p-hacking during his reign. We speculate that it was his emphasis on "facing reality" that held p-hacking in check at GE in the 1980s and 90s. A quick examination of Robert Slater's book, *Jack Welch and the GE Way*, written with Welch's input and cooperation, reveals that the word "reality" appears 76 times in 328 pages, or about once every four pages. The expression "face reality" or "facing reality" appears 27 times, including in a chapter title. Slater reports that Welch considered "facing reality" one of his most important business rules. With such a leadership emphasis, it would be difficult to move too far in the direction of the later stages of p-hacking, Losing Track of What's Real, Distracting from What's Real, and Artificial Reality.

GE's system of performance abstractions remained after Welch's retirement. But his emphasis on facing reality did not. Immelt, Welch's successor, was criticized for "success theater," a relentless optimism about the company's position and future, which too often turned out to be far from reality.[16] His sanguinity, some argued, created a culture in which people believed he didn't want to hear bad news[17] During his reign as CEO, the company's multi-business unit accounting practices were increasingly described as a "black box" used for "window-dressing" the company's

financial performance.[18] The Securities and Exchange Commission launched an investigation of the company's financial reporting in 2017, which was expanded in 2018.[19]

And, as we will recount, GE has in recent years seen its prowess with sophisticated management notions turn on it. Since about 2015, the company's stock has performed very badly, as the company seems to have fumbled an ambitious effort to transform the company into a major digital economy player. The vision of Jeffery Immelt, CEO from 2001 to 2017, though undeniably intriguing, has proven difficult to deliver, and the current CEO, Larry Culp, seems to be moving the company away from that plan. We will argue that part of the difficulty with the transformation was that performance assessment notions that served the company very well for a long time seem to have turned toward p-hacking. Analysts came to interpret the sophisticated multi business unit accounting that once seems so successful at GE as a means to artificially smoothing earnings.[20] Some of the company's difficulties getting customers to accept a new business model based on "selling outcomes," which required sharing of profits, revenues, or saving with clients according an agreed upon formula, might have appeared to clients too vulnerable to p-hacking – clever in concept, but too far removed from the delivery of operational outcomes to seem reasonable.

GE was arguably also the source of the p-hacking contagion at Boeing. The pivotal moment was the merger with McDonnell Douglas, which, because of the post-deal distribution of shares to individuals, left McDonnell Douglas executives and owners with a surprising amount of control over the combined company. Insiders described it as "McDonnell Douglas [buying] Boeing with Boeing's money." Executives from McDonnell Douglas, many of whom were GE alums, found their way into critical Boeing leadership positions. After an internal struggle, the GE culture, via McDonnell Douglas, prevailed at Boeing. The Boeing board hired GE alum James McNerney in

2005, as President and CEO. McNerney presided over the 787 rollout, and also made the fateful decision to go with an upgraded 737 MAX rather than build an entirely new airplane.

IBM – Stuck in Neutral

In the decades following the 1950s, the IBM System S/360 architecture became a dominant design that enabled entire industries to transform. IBM ruled the tech world and was viewed, in the words of former CEO Sam Palmisano, as "the greatest company in the world." In 1990, IBM was the second most profitable in the world. But in a sudden turn a year later, the company lost nearly $16 billion spread across nine subsequent quarters, and was on the verge of being sold off in pieces. CEO Lou Gerstner, who arrived in 1993, received deserved praise for saving the firm from this near-death experience. And yet in an industry that has exploded with growth since the mid-1990s, IBM has, since the early 90s turnaround, largely missed out on the party. The company's annual revenues in 1995, about $79 billion, are almost exactly the same 25 years later.

Much has happened in those 25 years. After the turnaround, Gerstner called for an embrace of the Internet (which IBM had earlier dismissed as "a university thing") and a new emphasis on strategic middleware, which led to key acquisitions (e.g., Lotus). Executives identified "Emerging Business Opportunities" (EBOs) and invested in them aggressively, making progress with some. The company's services business expanded dramatically; by 2000, services provided almost 40% of IBM's $88 billion in sales. But as Palmisano took the reins in 2002, the dot-com bust had largely erased IBM's hard-won progress. Palmisano doubled down on services, acquiring the consulting arm of Price Waterhouse Coopers and selling off the company's PC division. He pushed "on-demand" computing, cloud computing, and Software as a Service (SaaS), driving sales up to a peak $105 billion in 2011.

But when Ginni Rometty took over in 2012, she inherited daunting challenges. Legacy businesses were shrinking faster than the company was replacing their revenues. She looked for new growth and ultimately bet on *Watson*, IBM's AI technology, which had become famous by besting human players in the TV game show Jeopardy. The company organized around Watson, marketing world-best analytics capabilities delivered in the form of Watson-as-a-Service (WaaS). The company's $79 billion in revenues in 2018 reflect divestitures that are part of Rometty's strategy, which, admittedly, could still play out favorably.

Nevertheless, it is difficult to evade the observation that rivals over this same period have exploited growth possibilities much more successfully. From 1995 to 2018, SAP – which has dominated a large business systems software market that might have belonged to IBM – grew from $4 billion to $30 billion. Microsoft grew from $6 billion in 1995 to $110 billion in 2018. After having its own near-death experience in the early 1990s, Oracle, then a primary IBM rival in database management software, expanded from $2 billion in 1995 to almost $40 billion by 2018. Comparisons with newer firms are even more striking; Amazon, now the leading web services provider, made only $500,000 in revenues in 1995 (its second year of existence), compared to $26 billion in web services alone, with a current annual growth rate near 50%. Google did not exist in 1995 (it was founded in 1998), but this major rival to IBM's AI supremacy achieved $136 billion in revenues in 2018 (or, rather, its parent company Alphabet did).

We suggest a p-hacking explanation of IBM's inability to grow within an industry exploding with growth. The early 1990s crisis at IBM was borne of a particularly explicit p-hacking scheme, in which selling formerly leased hardware, moving it off the balance sheet by cashing out assets, masked slipping sales in the mainstream business. And while Gerstner did indeed make substantive changes at IBM, his stock buybacks and other schemes to increase the

company's earnings per share ultimately laid the groundwork for manipulating appearances rather than adapting to real changes. Palmisano subsequently embraced earnings management. And Rometty attempted to pass off the results of a television gameshow as a basis for corporate strategy, to ultimately similar effect.

What P-Hacking Companies Lose and the P-Hacking Pattern

If p-hacking were purely a matter of questionable behavior, then it might be possible to disapprove of it without worrying about the consequences. But we believe p-hacking does have consequences, major ones. Here are two big ones:

1. *Unrealistic piloting leads to a crash.* Just as an airplane will crash if a pilot stops guiding it in a way that takes reality-based hypotheses about of performance seriously, so a company will eventually have major problems when its leaders start to favour of questionable hypotheses and interpretations about the performance of a company. The most obvious recent example in our set of three companies is Boeing, but GE is not far behind.

2. *It's impossible to fly high if you insist on always flying smooth and level.* A company that has fully moved to to a risk avoidance regime, and progressed unchecked to later stages of p-hacking, can no longer construct major new value creation platforms. Boeing could not make the numbers come out in favour of building a new airplane, of betting the company, as it has done repeatedly throughout its history. They chose instead, with the 737 Max, to modify an old airplane, ironically avoiding more the appearance of risk than actual risk, which has since materialized with a vengeance. GE lost the ability to execute Immelt's vision to build a tech platform that could complete with the likes of Amazon and Google. IBM lost its capacity

reclaim the kind of success that it saw with the IBM360, despite participating in one of the world's most explosively growing industry.

All three of these companies are great companies – or have been. All three arose and grew based on technological innovations (airplanes, computers, electricity and industrial equipment). All three evolved and were influenced by remarkable CEO leaders (e.g., Bill Boeing and Allen at Boeing; Coffin and Welch at GE; Thomas Watson at IBM) who established or maintained strong corporate cultures that were critically important to the corporations' growth and longevity. All were able to survive beyond the reign of their founders to achieve even more greatly.

But they are also companies whose executives somewhere along the line succumbed to bad habits. In some cases, these leaders addressed real problems and created real solutions, but without giving up the p-hacking habit, which later came back to haunt them (or their successors). Often, it starts with good intentions and good ideas. But eventually, the balance is lost and an accelerating descent down a slippery slope gets underway.

There is, we suggest, a p-hacking pattern. The first part in the pattern is the generation and reliance on performance abstations, which decompose corporate performance into facets. These facets reveal insights, which can help manage the company. But no one of them captures actual performance accurately. The second part of the pattern is when the abstractions that create the possibility of p-hacking co-exist with a healthy inherited culture, under the influence of a strong, reality-oriented leader (e.g., Welch). The potential for p-hacking exists in this situation, but the company does not progress to later stages of p-hacking because the strong leader's influence counteracts it and causes the culture to hold fast. The final part of the pattern happens when the strong leader steps aside and is replaced by leader with no commitment to the historical culture,

and an inclination to "play with the numbers" to get desired results. The culture starts to erode. Eventually, if p-hacking behaviours are not brought back into check, a crash becomes inevitable.

The Authors

Rob Austin and Dick Nolan are professor at Ivey Business School and Harvard Business School, respectively. They are the authors of several books, including the recent *Adventures of an IT Leader*, now in its second edition, which they authored jointly (Harvard Business Review Press, 2016, with Shannon O'Donnell).

Professors Austin and Nolan wrote the now classic, still best-selling Harvard Business School case on Gerstner's turnaround of IBM (2000), a project that involved direct interaction with the company's executives. Professor Nolan began his career at Boeing, has advised Boeing executives for decades, and is the author of an HBS case on Boeing, "Boeing 787: The Dreamliner" (2005), as well as a book about the 787 project, *Executive Team Leadership in the Global Economic and Competitive Environment* (Routledge, 2015). Professor Austin is the author of a recent Ivey Business School case "Digital Transformation at GE: What Went Wrong? (2019). His research on dysfunctional performance measurement, based on an award winning doctoral dissertation, and which provides some of the underpinning of our analysis of p-hacking, was summarized in a 1996 book, *Measuring and Managing Performance in Organizations.*

---

[1] The Aviation Safety Network Database, entries for the do Haviland DH-106 Comet, https://aviation-safety.net/database/types/de-Havilland-DH-106-Comet/losses, accessed on July 4, 2020.

[2] Erik Lacitis, "60 Years Ago: The Famous Boeing 707 Prototype Barrel Roll Over Lake Washington, *The Seattle Times,* August 7, 2015, https://www.seattletimes.com/seattle-news/60-years-ago-the-famous-boeing-707-barrel-roll-over-lake-washington/, accessed on January 18, 2020.

[3] Mike Lombardi, "JET POWER: The Dash 80 marked a revolution in aviation—and represents a philosophy that's still being followed," *Boeing Frontiers*, Vol. 3, Issue 3, July 2014,
 http://www.boeing.com/news/frontiers/archive/2004/july/ts_sf2.html, accessed on January 19, 2020.
[4] Don Duncan, "Barrel Roll in a 707? History-Making Pilot Ends Silence," *The Seattle Times*, February 28, 1990, https://archive.seattletimes.com/archive/?date=19900228&slug=1058594, accessed on January 18, 2020.
[5] Tex Johnston, 2014, *Jet-Age Test Pilot*, Smithsonian Books: Washington DC.
[6] Duncan, 1998.
[7] Lacitis, 2015.
[8] See, for example, Price, Richard A. Sharp, Nathan Y, and David A. Wood, "Detecting and Predicting Accounting Irregularities: A Comparison of Commercial and Academic Risk Measures *Accounting Horizons* (2011) 25 (4): 755–780.
[9] Richard L. Nolan, 2015, *Executive Team Leadership in the Global Economic and Competitive Environment*, Routledge: NY.
[10] Nolan, 2015.

[11] Jon Ostrower, "Boeing's Unique Accounting Method Helps Improve Profit Picture," *Wall Street Journal*, October 4, 2016, https://www.wsj.com/articles/boeings-unique-accounting-method-helps-improve-profit-picture-1475522362, accessed April 2017.

[12] Jonas Hess, Suraj Srinivasan, David lane, and James Barnett, "Accounting Turbulence at Boeing," Harvard Business School case #118-020, September 28, 2018.
[13] Frank T. Rothaermel and Christopher K. Zahrt, "GE after GE Capital," McGraw Hill Education case MH0033, May 19, 2015..
[14] John Winthrop Hammond, 2012, *Men and Volts: The Story of General Electric*, Forgotten Books, quoted in Rothaermel and Zahart, 2015.
[15] Jane Nicholls, "Inside Crotonville: GE's Corporate Vault Unlocked, GE Reports, GE website, October 29, 2017, https://www.ge.com/reports/inside-crotonville-ges-corporate-vault-unlocked/, accessed on January 19, 2020.

[16] Thomas Gryta and Ted Mann, "GE Powered the American Century—Then It Burned Out," *Wall Street Journal*, December 14, 2018, accessed April 19, 2019, www.wsj.com/articles/ge-powered-the-american-centurythen-it-burned-out-11544796010.

[17] Thomas Gryta, Joann S. Lublin, and David Benoit, "How Jeffrey Immelt's 'Success Theater' Masked the Rot at GE," *Wall Street Journal*, February 21, 2018, accessed April 16, 2019, www.wsj.com/articles/how-jeffrey-immelts-success-theater-masked-the-rot-at-ge-1519231067?mod=article_inline

[18] Geoffrey Colvin and Katie Benner, "GE under Siege," *Fortune*, October 15, 2008, accessed April 16, 2019, http://archive.fortune.com/2008/10/09/news/companies/colvin_ge.fortune/index.htm.

[19] Michael Sheetz, "GE Shares Fall as SEC and DOJ Expand Investigations of Accounting Practices," CNBC, October 30, 2018, accessed April 16, 2019, www.cnbc.com/2018/10/30/ge-says-sec-expanding-scope-of-ongoing-accounting-investigation-shares-fall.html; United States District Court Southern District of New York, "Case 1:18-cv-04746-UA," 11, May 30, 2018, accessed April 26, 2019, www.courthousenews.com/wp-content/uploads/2018/06/Generalelectric.pdf; Thomas Gryta, "SEC Has Opened Probe of GE's Accounting," *Wall Street Journal*, January 24, 2018, accessed April 16,

2019, www.wsj.com/articles/ge-shows-long-road-ahead-in-restructuring-push-1516797761?mod=article_inline; "GE Completes Acquisition of Alstom's Power and Grid Businesses," General Electric, press release, November 2, 2015, accessed April 19, 2019, www.ge.com/power/about/alstom-acquisition; General Electric, *FQ3 2018 Earnings Call Transcripts*, 2018, 5, accessed January 22, 2019, www.nasdaq.com/aspx/call-transcript.aspx?StoryId=4215813&Title=general-electric-ge-q3-2018-results-earnings-call-transcript; Michael Rapoport, "How GE Built Up and Wrote Down $22 Billion in Assets," *Wall Street Journal*, March 13, 2019, accessed April 16, 2019, www.wsj.com/articles/how-ge-built-up-and-wrote-down-22-billion-in-assets-11552469401?ns=prod/accounts-wsj.

[20] Geoffrey Colvin and Katie Benner, "GE under Siege," Fortune, October 15, 2008, accessed April 16, 2019, http://archive.fortune.com/2008/10/09/news/companies/colvin_ge.fortune/index.htm.