Rational Accidents

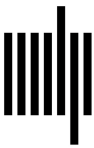
Reckoning with Catastrophic Technologies

By: John Downer

Citation:

Rational Accidents: Reckoning with Catastrophic Technologies
By: John Downer
DOI: 10.7551/mitpress/8844.001.0001
ISBN (electronic): 9780262377010
Publisher: The MIT Press
Published: 2024

The open access edition of this book was made possible by generous funding and support from MIT Press Direct to Open



The MIT Press

13 THE MYTH OF MASTERY: ON THE UNDERAPPRECIATED LIMITS OF TECHNOLOGICAL AMBITION

> The history of past investigation demonstrates that greater prudence is needed rather than greater skill. Only a madman would propose taking greater risks than the great constructors of earlier times.
>
> —Wilbur Wright

> Beware of false knowledge; it is more dangerous than ignorance.
>
> —George Bernard Shaw

13.1 FALSE PROMISES

TECHNOLOGICAL MASTERY

When America's least loquacious man stepped onto the Moon in 1969, most of his spectators appreciated the extraordinary dangers involved in getting him there. NASA had spoken repeatedly about its efforts to ensure Neil Armstrong's safety, but it had also been relatively forthright about the risks he faced. A great many would have been saddened if the *Apollo* 11 astronauts had not returned, but few would have been terribly surprised.

It is doubtful whether the assurances and expectations that accompanied the first Moon landing would have been quite the same if it had occurred today. Publics now expect organizations like NASA to promise more from complex technologies, especially in matters pertaining to safety. Many factors have contributed to this shift in expectations, but it has undoubtedly been facilitated by an increasingly caricatured portrayal of engineering safety

assessments as quantitative, objective, and definitive. By promulgating the idea that experts can interrogate technical systems with deductive certainty, the routine white-boxing of complex technologies in public discourse has made possible, and even encouraged, a corresponding belief that experts can make firm promises about the failure behaviors of complex systems. This belief underpins a deep-rooted sense of technological mastery—a widespread and institutionalized conviction that engineers can, and therefore *should*, know the reliability of the machines they build.

On one level, the premises underpinning this sense of mastery are fundamentally misleading—experts cannot perfectly interrogate their machines—but as we saw in chapter 2, the conceit is broadly functional in most contexts. The levels of reliability that experts claim of most technologies are not high enough for the epistemological limits of proof to make a material difference to those claims' usefulness. (It being much more practical, when designing, testing, or modeling systems, to anticipate circumstances that might occur every 10,000 hours, than to anticipate those that might occur every billion hours, especially when those systems are relatively straightforward.) No technology is entirely immune to rational accidents—the hidden uncertainties of technical practice always harbor the potential for rare but unexpected failures—but such failures are too rare to matter in most contexts.[1] So it is that a belief that experts can speak decisively and authoritatively about reliability has long contributed to a practical relationship between engineering and civil society. We might come to regret our technological creations for any number of reasons—the pollution they cause, the disruptions they bring, the costs they incur—but those creations usually function about as reliably as experts promise. And if they don't, it is usually appropriate to impute some level of avoidable error or shortcoming to the various experts responsible for them. Broadly, our intuitions, structures, practices, laws, and assumptions are all well adjusted to navigating most matters related to the probability of technological failure.

It is only in the context of what I have called catastrophic technologies—where the viability of complex systems depends on them being ultrareliable—that the inherent indeterminacies of knowledge threaten to meaningfully undermine the authority of expert claims about reliability. This is because, as we have seen, the fundamental uncertainties of engineering knowledge become uniquely consequential when making assurances about billions of hours of failure-free operation. In these circumstances, even extremely rare

accidents can dramatically shape reliability calculations, so even the most trivial uncertainties—born, perhaps, of marginal misjudgments about the equivalence between test conditions and real-world environments—become determinative. Here, therefore, where engineers work in the shadow of disaster and must establish extreme reliabilities in advance of service data, finitist dilemmas like the problem of relevance have a real bearing on the viability of their ambitions and the credibility of their assertions. In this context, if in no other, the idea of engineering mastery over reliability becomes misleading and potentially dangerous. No amount of formal analysis—testing and modeling—can realistically guarantee that a complex system will operate for billions of hours without encountering rational accidents.

UNDERAPPRECIATED

The unique difficulties that arise when complex technologies demand extreme reliabilities are important to recognize because they are systematically underappreciated. Reliability is the sine qua non of catastrophic technologies (almost by definition), but the way we speak about it and govern it in catastrophic-technological domains rarely, if ever, reflects the unique epistemological challenges that it poses.

There are several reasons for this, each of which has been outlined in preceding chapters. First and most fundamental, it is because the distinctiveness of engineering's relationship to catastrophic technologies is counterintuitive and nonobvious. As discussed previously, most of the claims that reputable engineering bodies make about most technologies are manifestly and deservedly credible, including most of those made about technological reliability. When respected experts make assertions about the properties of a system—be it "this engine will burn *x* amount of fuel per minute" or "this building will not be structurally sound"—it is wise to listen. The claims that they make about ultrahigh reliability in catastrophic technologies are an exception to this rule. These are much less credible than most other engineering claims, as we have seen, but it is far from obvious why this would be the case. Our heuristics about the authority of engineering expertise fail us in this context.

Second, it is because very few catastrophic technologies accrue enough service data for any deficit in their reliability to become self-evident. Any technology that is built and operated in limited numbers—tens or even hundreds of units—would still fail incredibly infrequently, even if its mean-time-to-failure was 100 times, or even 1,000 times, less than we require of catastrophic

technologies. Where a dearth of service experience makes extreme reliability impossible to fully master, in other words, it simultaneously hides that lack of mastery from the public. Where catastrophes do still occur in these circumstances, they do so with such rarity that they can invariably be explained away, blamed on specific design weaknesses or organizational shortcomings instead of being understood as evidence of fundamentally intractable problems (Downer 2014; Hilgartner 2007).[2] Ultimately, humans are predisposed to think at human scales (Kahneman 2011), so any system that operates failure free for decades intuitively feels like evidence of ultrahigh reliability, even if its required mean-time-to-failure is north of a million years. (Hence the absence of civilization-ending atomic wars is routinely cited as evidence that deterrence networks have obviously made us safer.)

A third reason why we underappreciate the difficulties of making catastrophic technologies ultrareliable is that the only such technologies that we operate at enough scale for their reliability to be statistically (and intuitively) visible are jetliners, and jetliners, confoundingly, have proved to be as reliable as experts have promised. So the fact that reliability in jetliners is ostensibly governed in the same way as in other catastrophic technologies—rigorously interrogated via tests and models, under the watchful gaze of a dedicated regulator—makes their remarkable (and highly contingent) failure performance look misleadingly generalizable. If jetliners were crashing 100 or 1,000 times more frequently, then it is reasonable to assume that publics and policymakers would more intuitively appreciate the limits of engineering knowledge and the implications of those limits for technologies like reactors and deterrence networks. With this final consideration in mind, therefore, it is worth pausing to recap why it is, exactly, that the reliability of jetliners is so misleading.

FALSE EQUIVALENCE

As we have seen at length, the extreme reliability of jetliners is misleading because it was not achieved in the positivist manner that we are led to believe via formal analysis and rule-governed oversight. Airframers evolved that reliability gradually, by exploiting real-world experience on a massive scale to whittle the uncertainties of an uncommonly stable design paradigm. Regulators honed their assessments in much the same fashion, learning from hard experience how frequently jetliners built around that paradigm failed in real-world operations. In both cases, the process took decades,

exacting an enormous cost in accidents and lives. It was practicable only because jetliners were built and operated in far larger numbers, and with far deeper commonalities, than other catastrophic technologies. These commonalities were only sustainable organizationally because civil aviation's operating volume—among other conditions—gave rise to an uncommon incentive structure, wherein perceptions of unreliability were heavily and quickly punished financially. And it was endurable politically only because public risk tolerances evolved with the industry itself, a process that was possible only because jetliners are chronic catastrophic technologies; their extreme reliability demands being driven as much by their volume as by the catastrophic potential of any individual failures.

To be more methodical, the preceding chapters outlined four conditions on which the process outlined depends:

1. *A long legacy of expansive service from which to learn and on which to build*: billions of cumulative operational hours, born of decades of flight by tens of thousands of jetliners with substantial design commonalities, that created a deep well of experience, including a slew of catastrophic failures, which experts could mine extensively for the marginal, esoteric insights required to hone their designs and assessments.
2. *A longstanding commitment to recursive practice*: an institutionalized willingness to invest heavily in the difficult and expensive work of mining the service experience in the manner outlined here, exploring and generalizing from failures and near-misses.
3. *A longstanding commitment to design stability and innovative restraint*: an institutionalized willingness to forgo the promises of radical innovation to aggregate the lessons of the past, manifest as a commitment to a common design paradigm that airframers deviate from rarely, incrementally and extremely carefully.
4. *Structural economic incentives that support these longstanding commitments*: a set of circumstances—such as the existence of competition in the industry, and a tendency of passengers to punish failures in their purchasing decisions—which, in combination with civil aviation's expansive service, tend to quickly punish any systemic unreliability.

These are all *necessary* (albeit, again, not *sufficient*)[3] conditions. Remove even one, and the challenges of ultrahigh reliability become prohibitive. And they all have significant interdependencies. The design stability contributes

to the accrual of relevant service experience, for example, and makes the lessons learned from it more useful, for example, while the service experience is crucial to the incentive structure and provides the lessons from which to learn.

It is consequential, therefore, that no other catastrophic technology enjoys the same confluence of conditions. From deterrence infrastructures to drilling platforms, reactors to financial networks, few enjoy even one of these conditions to the same degree as jetliners. (This shouldn't be surprising, given that their interdependencies make each more difficult to achieve without the others.) None are designed with an equivalent commitment to design stability or are operated in equivalent numbers, so none accrue service experience at the same rate as jetliners. This dearth of service experience, in turn, means that the experts who manage and build these technologies have far fewer accidents to mine for the subtle but dangerous misunderstandings that lurk in the epistemology of their designs and assessments. It also means that any insights that those experts do glean from service are often rendered moot by changing designs. And, beyond that, it means that the incentive structures that frame their work are unlikely to consistently favor ultrahigh reliability, with the many costs and sacrifices it demands. (The dearth of service experience allowing reliability shortfalls to remain hidden for long periods.)

Most catastrophic technologies could not enjoy these conditions even if societies were committed to re-creating them. There is no world in which we could build and operate as many reactors or ultradeep drilling platforms as there have been jetliners, and no world in which we could let them fail in the same way. Chapter 1 outlined two types of catastrophic technology, each requiring comparable levels of failure performance, but for slightly different reasons. These were (1) chronic catastrophic technologies, which can tolerably be allowed to fail on rare occasions but require ultrahigh reliability because of the volume at which they operate; and (2) acute catastrophic technologies, which operate at much smaller volumes but require ultrahigh reliabilities because their failures are wholly intolerable. Jetliners are chronic catastrophic technologies: we operate them in large numbers and broadly tolerate very infrequent accidents. Most other catastrophic technologies—deterrence networks, reactors, financial instruments, or drilling platforms—are acute. We build them in much smaller numbers but with comparable reliability demands because they can *never* be allowed to catastrophically fail (as the consequences of such failures could be truly intolerable). There can

be no trial-and-error learning with such systems; no "searching for safety," as Wildavsky (1988) once put it.

So it is that civil aviation's technique of ratcheting systems to ultrahigh reliabilities is simply not viable in most other catastrophic-technological domains. In these domains, therefore, experts pursuing extreme reliabilities must *actually* operate in the positivist manner that those in civil aviation only purport to operate, with all the epistemological perils and limitations this implies. This is to say that their designs, and any assessments of their designs, have to be wholly grounded in knowledge gleaned from tests and models. And any shortcomings in those tests and models—born perhaps of imperfect relevance assumptions—are liable to become shortcomings in their designs and assessments, each a potential rational accident. The organizations that structure this work also have to operate against their own economic incentives, which rarely reward the kind of expenditures and sacrifices that extreme reliability requires.

Bluntly put, the reliability achieved in jetliners is unlikely to be achievable elsewhere. Civil aviation's service record demonstrates that modern societies are, in principle, capable of building a complex sociotechnical system that predictably operates for billions of hours between fatal surprises. What it does *not* demonstrate, however, is that this achievement is necessarily transferrable. Jetliners are evidence that experts have mastered extreme reliability in a very specific design paradigm, not that experts have mastered the tools and processes for achieving ultrahigh reliabilities more generally. In this regard at least, most catastrophic technologies are in no way equivalent to jetliners, and the reasons we trust the latter should give us cause to doubt the former.

The most serious harm that arises from white-boxing jetliner reliability is that it hides this difference. Presenting civil aviation's record as grounded in objective, positivist processes hides the informal foundations—conditions and commitments such as design stability, service data, and recursive practice—on which jetliner reliability is built; and masks the price at which that reliability was bought. In doing so, it obscures the uniqueness of civil aviation, and with it the costs and difficulties of replicating its reliability achievements elsewhere. It gives us no reason to doubt the efficacy of tests and models, no reason to look too closely at structural incentives, and no reason to imagine that extreme reliabilities *need* to be incubated over long periods of unreliability. Instead, it suggests that any complex system, if designed, governed, and assessed with equivalent quantitative rigor, might perform as

reliably as a modern jetliner (and expert assertions about that system's reliability might be as credible).

The fact that the reliability issues pertaining to all catastrophic technologies are portrayed in the same way in all catastrophic-technological domains, and are ostensibly managed in the same fashion—via rules and standards that call for careful measurements and are enforced by objective regulators—colludes in this illusion. So does the limited service experience that most catastrophic technologies accrue: an apparent absence of accidents (born of sparse service) making it easy to believe (and difficult to disprove) that unrealistic reliabilities have been achieved.

By attributing the reliability of jetliners to structures and practices that look equivalent across domains, the white-boxing of aviation safety invites us to use aviation as a touchstone by which to gauge our mastery of catastrophic technologies. In doing so, it creates a dangerous misapprehension. It is an irony of modernity that the only catastrophic technology with which we have real experience, the jetliner, is highly unrepresentative, and yet it reifies a misleading perception of mastery over catastrophic technologies in general. Meanwhile, that misleading perception of mastery itself helps hide the unrepresentativeness of jetliners.

13.2 THE PERILS OF CERTAINTY

TECHNOLOGIES OF HUBRIS

The widespread misapprehensions outlined in the previous section have made modern societies poor at reckoning with catastrophic technologies. Unlike most other expert engineering assertions, those made about the reliability—and thus the safety and viability—of our most dangerous creations are simply not credible. This is hidden, however, by an idealized understanding of engineering knowledge, deeply entrenched in our institutional logics: an understanding implying that experts speak about extreme reliability with the same authority as they speak about other engineering variables, and that they can achieve such reliabilities in complex systems without gradually whittling them from painful service experience.

This misplaced sense of mastery, with its implausible promise of extreme reliability, is dangerous. There is an old adage that "great things are achieved by those who don't know that failure is inevitable," and while this may be

true, the same cannot be said of prudent policymaking (or insightful scholarship) with regard to technologies that cannot be allowed to fail.

The dangers of this promise are accentuated by its reach and invisibility. It permeates discourses and decision-making around catastrophic technologies; explicitly—or, more often, implicitly—underpinning a range of consequential academic, legal, strategic, administrative, and legislative edifices.

In academia, for example, it is heavily implicated in various influential studies, from historians, sociologists, psychologists, and others, that treat expert assertions about the safety of catastrophic technologies as incontrovertible facts, to then be contrasted with (seemingly irrational) public perceptions of risk (e.g., Slovic 2012; Weart 1988; Douglas and Wildavsky 1982; Erikson 1991; Sjöberg 2004; Taebi 2017; Starr 1969). It is similarly implicated in analyses comparing the economic or environmental costs of various energy options, which almost never consider the possibility of reactor failures; in security scholarship that strategizes nuclear weapons deployment with little thought about potential accidents; and in myriad other judgments across a wide range of disciplines.

More materially, and arguably more consequentially, however, this promise is manifest in the decisions that such scholarship informs; the strategic choices societies make about what they build and how they build it.

Consider, for example, recent trends in petroleum extraction. A decade after the loss of Deepwater Horizon, with its extraordinary harms and associated costs—over 200 million gallons spilled and tens of billions of dollars in cleanup costs alone—ultradeepwater drilling platforms have continued to proliferate. By 2017, they were producing 52 percent of all oil in the Gulf of Mexico, up from just 15 percent in the decade leading up to the disaster (Murawski et al. 2020). Oil companies are deploying these platforms in ever-deeper waters—up to twice the depth at which Deepwater Horizon operated—and are drilling ever deeper into the ground, where temperatures and pressures are even more extreme. The backdrop of these trends is an attitude, said to be pervasive within the industry, that evolved quickly from "An accident like Deepwater could never happen," to "An accident like Deepwater could never happen again" (Calma 2020).

Consider also the recent burgeoning of "synthetic credit products": complex financial technologies like the collateralized debt obligations (CDOs) and credit default swaps that failed in 2007–2008, almost bringing the entire

global financial system to its knees (Gorton 2012; King 2016). For all the public recrimination that that crisis prompted, there is little evidence that societies have learned to be appropriately wary of the innovations that allowed it to occur. Trade in CDOs—which Warren Buffett presciently referred to as "financial weapons of mass destruction" in a 2002 shareholder letter (Graffeo 2021)—quickly resumed after a postcrisis lull (Boston 2019). Perhaps more ominously still, CDOs have since been eclipsed by another synthetic credit product called "collateralized loan obligations (CLOs)." Similar to CDOs but built on loans made to (often highly troubled) businesses rather than homeowners, CLOs have equal catastrophic potential, if not more (Partnoy 2020). The 2007 market for CDOs is estimated to have been around $640 billion; the 2020 market for CLOs was estimated to be over $870 billion (Partnoy 2020). Should a large number of risky businesses fail at the same time due to an unforeseen common-cause failure—another pandemic, for instance—then this market could collapse, again exposing the banks. Yet there has been minimal preparation for any such catastrophe. Banks deemed "too big to fail" in 2008 have grown even larger in the intervening years. (J. P. Morgan, for example, doubled in size between 2008 and 2018.) And regulations put in place to improve the industry's resilience have started to be rolled back (Li 2018).

Beyond drilling and banking, an even more compelling illustration of the promise of implausible reliability at work is offered by atomic weapons—arguably the original catastrophic technology. The strategic decision-making around such weapons rarely, if ever, grapples meaningfully with the possibility of catastrophic failure; a convention that has persisted through a litany of close calls with disasters that almost defy comprehension (Sagan 1993; Schlosser 2013).

These close calls include a slew of chilling incidents involving individual warheads. In 1961, for example, a technological failure led the US Air Force to accidentally drop two hydrogen bombs on Greensboro, North Carolina. One was only kept from detonating by a small, damaged switch, one of four redundant safety measures, which a recently declassified investigation claims might easily have shorted (Pilkington 2013; Schlosser 2013, 246).[4] A similarly alarming event occurred in 1980, at a Titan II missile silo in Damascus, Arkansas, when a dropped wrench led to an explosion that catapulted a 740-ton silo door into the air, together with a 9-megaton nuclear warhead (Schlosser 2013, 392–398). The warhead—said to have been three times more powerful

than every bomb dropped during World War II combined—is again now thought to have been at meaningful risk of detonation (Scholsser 2013, 440; O'Hehir 2016). These are not isolated examples. Historians are uncovering a chilling number of potentially catastrophic malfunctions and mishaps with atomic weapons as the mists of secrecy slowly lift on the Cold War. Schlosser (2013) reports close to 1,200 dangerous events occurring just between the years 1950 and 1968.

Even more alarming, insofar as that is possible, are a slew of near-misses arising from technological failures in the deterrence system itself—failures that might have caused atomic wars rather than isolated explosions. Again, these events are shrouded in a veil of secrecy, such that historians know a lot more about the earlier years of the Cold War about later years. In that period alone, however, they have identified a slew of occasions where false alarms about Soviet missile launches—many caused by technological failures—mobilized the fast-moving machinery of US nuclear retaliation, only for it to be stood down before any commitment became irrevocable. In October 1960, for example, an early-warning radar system in Greenland misinterpreted the Moon as a major incoming Soviet missile strike (Stevens and Mele 2018; UCS 2015). (It created a panic that could have been significantly worse if Soviet premier Nikita Khrushchev hadn't serendipitously been in New York at the time.) In November of the following year, a failed relay station led US Strategic Air Command Headquarters to lose contact with the North American Air Defense Command (NORAD) and multiple early warning radar sites simultaneously, a situation only thought possible in the event of a coordinated attack (UCS 2015). In November 1965, a mass power outage combined with a series of malfunctions in bomb-detecting equipment created a compelling illusion of nuclear attack, prompting another major alert (Philips 1998). In May 1967, radar interference from a solar flare was interpreted as intentional Soviet jamming intended to cover for a nuclear attack, almost leading to a counterstrike (Wall 2016). Jumping forward to November 1979, a computer error at NORAD headquarters led the agency to inform the US national security advisor that the Soviet Union had launched 250 missiles (a figure that it subsequently revised to 2,200) at the US, and to announce that it needed a decision on retaliation in three to seven minutes (Philips 1998; Stevens and Mele 2018).

By general consent, however, the most dangerous such incident known to historians came from the Soviet side. On September 26, 1983, one of the

Soviet Union's early warning systems mistook sunlight reflected off clouds as a US missile launch. The malfunction came at a time of significant tension and distrust between the Soviet Union and the US, and many believe that, had the duty officer at the radar station, Stanislav Petrov (whose name deserves to be remembered), reported the launch as protocol dictated, then a retaliatory launch (and ensuing thermonuclear war) would almost inevitably have followed. Civilization rolled two sixes, and Petrov, who distrusted the technology, chose to defy protocol and wait for further confirmation (Myre 2017; Schlosser 2013, 447–448). As General Lee Butler, a former head of US Strategic Command, put it, "we escaped the Cold War without a nuclear holocaust by some combination of skill, luck and divine intervention—probably the latter in greatest proportion, . . . [b]ecause skill and luck certainly don't account for it" (quoted in Kazel 2015).

This is all to say that we have long been tempting fate, and the age of catastrophic technologies is only in its infancy. As Beck (1992; 1999) and others have observed, decades of near-misses and painful lessons have not meaningfully lessened the confidence with which we confront the risk of technological failure. A host of new catastrophic technologies are emerging without eliciting appropriate levels of concern in this regard. Entranced by the promise of implausible reliability, and implausible certainty about that reliability, our appetite for innovation has outpaced our insight and humility.

The technologies discussed here speak to this dynamic. The fact that sunlight reflecting off clouds could have instigated the end of the world as we know it probably should weigh more heavily on our understanding of progress. But nothing exemplifies and illuminates the issue better and more clearly than the story of nuclear reactors.

By way of a coda, therefore, let us return, finally, to Fukushima.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Stone Sans and Stone Serif by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Downer, John (John R.), author.
Title: Rational accidents : reckoning with catastrophic technologies / John Downer.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Series: Inside technology | Includes bibliographical references and index.
Identifiers: LCCN 2023002845 (print) | LCCN 2023002846 (ebook) | ISBN 9780262546997 (paperback) | ISBN 9780262377027 (epub) | ISBN 9780262377010 (pdf)
Subjects: LCSH: Reliability (Engineering) | Aircraft accidents—Prevention. | Risk assessment. | Industrial accidents—Prevention.
Classification: LCC TA169 .D69 2023 (print) | LCC TA169 (ebook) | DDC 620/.00452—dc23/eng/20230202
LC record available at https://lccn.loc.gov/2023002845
LC ebook record available at https://lccn.loc.gov/2023002846