Rational Accidents

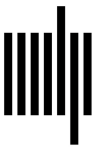
Reckoning with Catastrophic Technologies

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

4 ORGANIZING AVIATION SAFETY: RELIABILITY REQUIREMENTS AND LOGICS

When we say an airline is safe to fly, it is safe to fly. There is no gray area.
—David Hinson, FAA Administrator, 1996

Recently a man asked whether the business of flying ever could be regulated by rules and statutes. I doubt it.
—Walter Hinton, 1926

4.1 RULES AND INSTITUTIONS

TYPE-CERTIFICATION

On witnessing a demonstration of the UK's first jet engine in January 1940, Winston Churchill's influential science advisor, Henry Tizard, is said to have defined a "production job" as any prototype that did not break down in his actual presence (Constant 1980, 192). In the years since then, aviation reliability requirements have become considerably more exacting. Today, the performance of civil airframes is subject to a rigorous accountability program, wherein manufacturers work closely with regulators to establish failure behaviors long into the future. The entity responsible for coordinating this work in the US is the Federal Aviation Administration (FAA). Arguably the most prominent regulator of a complex technology anywhere in the world, the FAA has many duties. Prominent among them, however, is a mandate to police the reliability of new airframe and engine designs prior to their operation.[1]

The FAA conducts this work out of three regional Aircraft Certification Offices[2] through a process that it calls "type certification." In theory, it performs this process in parallel with its European counterpart, the European Aviation Safety Authority (EASA), with both agencies issuing independent type certificates for each new airframe design. In practice, however, there is a fairly substantial division of labor, with the FAA taking the lead in certifying airframes built by US manufacturers (principally Boeing in this category), while EASA leads in certifying aircraft built by European manufacturers (principally Airbus). This cooperation is governed by bilateral airworthiness agreements and facilitated by the fact that both agencies work from standards that have been harmonized to the point of being almost identical in their wording. Together, the two agencies all but monopolize international aviation certification, with every other nation either formally or informally recognizing the type certificates of one or the other.

The extensive international deference and cooperation around type certification reflects the sheer scale of the process. Certifying a new airframe type is a forbiddingly onerous undertaking. It generates tons of paperwork,[3] takes years to complete, costs hundreds of millions of dollars, and involves skilled work by thousands of highly qualified people. For example, certifying the Boeing 777 took half a decade (1990–1995) and directly involved over 6,500 Boeing employees, as well as an unknown number of subcontractors and FAA personnel. The process utilized nine test airplanes, which collectively accumulated over 7,000 hours of flight time over the course of 4,900 test flights (NTSB 2006b, 75).

Type certification's myriad rules, protocols, and standards are codified in an extensive pyramid of guidance material, which stipulates, with ever-increasing specificity, technical requirements for the design and assessment of each element and system in a jetliner, from its structural beams to its belt buckles. At the top of this pyramid is Federal Aviation Regulation Part-25, "Airworthiness Standards: Transport Category Airplanes," commonly referred to as "FAR-25."[4] FAR-25 governs the overall design of jetliners as integrated systems. It is issued by the FAA directly, as is most of the guidance that sits directly below it (which primarily takes the form of Directives, intended for regulatory personnel and comprised of Orders and Notices, and Advisory Circulars.)[5] Many of the more specific, downstream materials to which these documents refer, however, are authored by a wide array of bodies, ranging

from parallel government agencies such as the NRC or the US Department of Defense (DoD)[6] to nongovernmental organizations like the Radio Technical Commission for Aeronautics (RTCA), a volunteer organization sponsored as a Federal Advisory Committee by the FAA. Together, this great body of rules, requirements, codes, standards, best practices, and other miscellaneous guidance represents a sprawling, labyrinthine metatext of esoteric terms and layered definitions (see, e.g., appendix A of NTSB [2006b]).

PREDICTIVE AND POSITIVIST

In keeping with wider conventions around technology certification, as discussed earlier, a noteworthy feature of the guidance that governs type certification, from FAR-25 downward, is that it is predictive and positivist, largely treating each system's future reliability as an objectively quantifiable variable. This wasn't always the case. When US authorities began certifying airplane designs in 1926,[7] their requirements did not purport to calculate, quantify, or otherwise measure the reliability of the machines that they governed. Instead they established a set of proscriptive design rules for manufacturers, codified in a handbook published by the US Department of Commerce. For example, these rules specified minimum load factors for wings and mandated certain cockpit instruments. Upon designing a new aircraft for commercial use (private aircraft were excluded), manufacturers would submit blueprints to a newly created regulator,[8] which would check the design's conformance to federal mandates. If the regulator deemed the blueprints to be satisfactory, it subjected a prototype to a series of flight tests, upon the successful completion of which it issued an approved type certificate authorizing the airplane's sale and operation (Briddon et al. 1974; Komons 1978, 99).

The type-certification process retained this essential character for several decades, with its design mandates growing increasingly detailed and voluminous. Beginning in the middle of the twentieth century, however, its essential nature began to shift. The emphasis on specific design prescriptions slowly gave way to a more calculative idiom, wherein rules were framed in terms of minimum quantitatively specified reliability requirements. This shift to minimum reliability requirements represented a subtle but meaningful change in the way experts imagined, performed, and communicated the logic of type certification. It altered the definition of a "certified" aircraft from "an aircraft shown to have satisfied every design requirement" to being "an aircraft

shown to have a specified failure performance." In doing so, it premised the whole process on the ability of experts to interrogate designs and accurately quantify their failure behavior.

There were many reasons for this transition, including the aforementioned reconceptualization of reliability (from virtue to variable) in engineering more broadly, together with wider structural incentives toward quantification (which will be outlined in chapter 12). Internally, however, the US aviation community usually explains the shift in one of two ways, which are distinct but not mutually exclusive. The first explanation construes it as a response to the growing complexity of airplanes. By this view, the FAA started calculating the reliability of systems because their designs were becoming too elaborate to judge qualitatively. "As the number, criticality, complexity, integration, and number of parts of aircraft systems increased," explains an FAA publication, so "the combinations of conditions and events that a design must safely accommodate became more difficult to effectively judge by qualitative means alone" (FAA 2002c, 23). The second explanation construes the shift as an attempt to better accommodate innovation in the industry. By this view, the need for manufacturers to meet proscriptive design requirements was constraining their ability to reimagine airplane systems and architectures. Framing regulations around minimum reliability requirements was seen as a way of mitigating this. In theory, at least, it allowed engineers to build airplanes however they chose, so long as they could demonstrate that their designs met a satisfactory level of failure performance.

The exact mechanisms by which manufacturers are supposed to demonstrate conformance with type certification's reliability requirements vary. (The process treats "systems" differently from "structures," for instance.)[9] In most instances, however, the procedures are flexible, such that regulators will consider evidence from a vertiginous range of analytical tools.[10] The competing logics of these tools are highly esoteric, but they all draw on and manipulate data derived from the same foundational processes.

Ostensibly at least, these processes are the same for jetliners as they are for reactors and most other catastrophic technologies. This is to say that when stripped to their barest fundamentals, they involve two basic steps. The first is *testing*, wherein experts use controlled environments to empirically examine the failure performance of a jetliner's individual elements. As we will see, however, even the most extensive and idealized testing regimen could never demonstrate the levels of reliability that jetliners require;

it would take a prohibitively long time (potentially thousands of years) and the cost would be astronomical. What cannot be demonstrated empirically, therefore, must be demonstrated in principle. Hence the second step: *modeling*, wherein experts integrate their test results into representations of the wider system architecture and invoke those representations to demonstrate much higher reliabilities than could be established with tests alone.

Subsequent chapters of this book will examine these processes and their applications in more detail. Before then, however, it is worth pausing to consider the nature of the reliability requirements themselves. Just as tools for interpreting reliability data can only be as good as the data they interpret, so procedures that use reliability metrics to govern airframes necessarily hinge on the validity of their requirements. It doesn't much matter if experts are failing to measure a system's performance accurately—in other words, if the level of performance they are trying to ensure is insufficient to achieve their required ends.

4.2 RELIABILITY TARGETS

TWO NUMBERS

Perhaps the most interesting aspect of the FAA's turn to quantitative reliability requirements is how imprecisely it articulates the levels of reliability that are needed. Broadly, the failure performance that the FAA requires of new jetliners is usually expressed in terms of one of two numbers: one in ten million (10^{-7}), and one in a billion (10^{-9}) (FAA 1982, 2002b, 2002c). These numbers—10^{-7} and 10^{-9}—frame the entire type-certification process and are essential to the positivist vision of safety promulgated by the FAA and other aviation bodies around the world (e.g., Lloyd and Tye 1982). Examined closely, however, their underlying rationales are ambiguous in ways, and to degrees, that shed light on the viability of legislating for extreme reliability.

Let us consider each number in turn.

ONE IN TEN MILLION (10^{-7}) The first number, 10^{-7}, arguably represents the most fundamental goal of type certification, which is to ensure that the probability of a serious accident is no greater than "one in every ten million hours" of operation (FAA 2002b, 5). This figure, then, is the reliability that is officially required of a jetliner as an integrated system, and its origins and logic are ostensibly straightforward. Formally established in 1982, it

was simply an expression of the frequency of accidents per departure at the time. As such, it represented an implicit understanding that this accident frequency needed to remain constant, if not decline, if civil aviation were to remain credible among publics and policymakers (FAA 2002c, 23). On close inspection, however, it is probably fair to say that the calculation by which the FAA arrived at this number was not the most rigorous or extensive ever performed by a federal regulatory agency.

Its rationale hinges on at least two highly questionable premises. The first is that only one in every ten accidents is caused by a reliability issue. The accident rate at the time when the number was established was actually calculated to be about one in every million hours (thus 10^{-6}), but since only one in every ten of those accidents was attributed to system failure, the probability of a reliability issue felling an airplane was calculated to be $10/10^{-6}$ or 10^{-7} (FAA 2002c, 23). This "one in ten" attribution is highly problematic, however. Of the remaining nine accidents, most were attributed to human error, and, as accident theorists have long maintained, the distinction between "human error" and "technological failure" is blurry at best (e.g., Reason 1990; Perrow 1983). ("There is no problem so complex that it cannot simply be blamed on the pilot," as an old industry adage puts it.)

The second questionable premise in the FAA's adoption of the 10^{-7} number is the assumption that a stable rate of accidents per departure would remain acceptable as the absolute number of flights (and thus the absolute number of accidents) increased dramatically. This is problematic because, as we have seen, public opinion on aviation safety appears to be more sensitive to the *absolute* frequency of accidents than it is to the *relative* frequency. The number of accidents per year tends to loom larger than the number of accidents per departure, in other words. And since the number of departures per year has risen steadily since 1984, it is highly doubtful whether maintaining the same number of accidents per departure (i.e., the 10^{-7} reliability target) would be deemed acceptable today, given that it would represent a very considerable spike from current accident levels. (In fact, the 10^{-7} number was already outdated in this regard even at the time it was adopted, having been lifted from an earlier British Civil Airworthiness Requirement [FAA 2002c]). Fortunately, the number of accidents per departure has dropped steadily since 1982, such that modern jetliners now far outperform the 10^{-7} requirement (Flight Safety Foundation 2018).

ONE IN A BILLION (10^{-9}): THE "NINE 9s" The 10^{-7} reliability target is problematic, therefore, but it is arguably less important to the certification process than the second number: 10^{-9}. This is because type certification's standards are not really framed around the jetliner as an integrated system so much as they are framed around its individual systems and assemblies. (As the NTSB [2006b, 90] puts it: "[A]irplane-level risk and hazard analyses are neither required by certification regulation nor recommended by FAA advisory materials.") Rather than assessing the reliability of a whole jetliner, in other words, regulators assess the reliability of its essential elements: the operational unit of the certification process being the "flight-critical system," (which I will henceforth refer to as "critical systems") (FAA 2003).[11] Critical systems are the subassemblies of a jetliner that would jeopardize the plane if they failed—the flight controls or landing gear, for instance[12]—and it is to them that the 10^{-9} figure refers. It is widely held that airframers must demonstrate for each critical system a mean-time-to-failure higher than a billion hours of operation.

This number, 10^{-9}—often informally referred to as the "nine 9s" since it can be expressed as a reliability of 0.999999999, where 1.0 would represent perfection—appears frequently in the discourse around type certification. By most accounts, it is the linchpin of the process, representing the key reliability metric that experts are striving to achieve in their designs and validate in their assessments. It should be somewhat surprising, therefore, that the FAA never explicitly defines the 10^{-9} reliability target. The literature around certification offers competing rationales for it, often in the same documents and sometimes even on the same page.

At the most fundamental level, these rationales derive the number from one, or often both, of two distinct certification requirements, each of which, again, rests on problematic assumptions. The first derives it directly from the requirement outlined previously: that airplanes should crash no more than once in every 10 million (10^{-7}) hours. This rationale assumes that every jetliner has exactly 100 critical systems, each of which therefore has to be 100 times more reliable than the reliability required of the airplane itself.[13] The essential math here is relatively simple. If there are 100 sources of potential failure that could cause an airplane to crash in any given hour, then the probability of any one of those faults occurring has to be 100 times lower than the reliability required of the airplane itself ($100/10^{-7} = 10^{-9}$). But deriving 10^{-9} from the 10^{-7} requirement in this way introduces hidden complications. As outlined earlier, for instance, the 10^{-7} requirement itself rests on ambiguous

foundations; and insofar as it inadequately expresses the reliability required of each jetliner, then 10^{-9} will inadequately express the performance required of its critical systems. It is also far from clear that modern airplanes in fact have exactly 100 critical systems, not least because the distinction between "critical" and "noncritical" systems is highly subjective. The FAA itself has called the distinction "necessarily qualitative" (FAA 1988, 7), and found that critical systems are "not consistently identified" (FAA 2002a). Others have argued that the distinction is almost meaningless, since almost any component or system can be critical in the right (or wrong) circumstances (e.g., Perrow 1999; Leveson et al. 2009; Macrae 2014, 8). (When a jetliner leaving Boston crashed in 1960, for example, investigators found that a faulty seat-adjustment pin had caused the pilot to make an "unintended input" [Newhouse 1982, 94–96]).

The second, more common rationale given for the nine 9s is derived from a requirement in FAR-25 that catastrophic failures be "extremely improbable," defined, elsewhere in the guidance, as meaning "not anticipated to occur during the entire operational life of all airplanes of one type [its fleet life]" (FAA 2002b, 9). By this rationale, the 10^{-9} reliability target is understood as the level of performance needed to satisfy this requirement. Again, however, deriving the number from the rule requires some questionable assumptions, not least because any calculation of the reliability required to make catastrophic failures "extremely improbable" by the definition given here necessarily hinges on both (1) how many planes of each type will enter service [i.e., the fleet], and (2) how long that service will last [i.e., the life]. Neither of these variables is defined in the guidance, and both have changed meaningfully over the last forty years.

Given this definitional ambiguity, it is probably unsurprising that when secondary literatures make more detailed attempts to justify the nine 9s, they tend to vary widely in their reasoning. Such efforts usually combine the two rationales: constructing an explanation that defines "fleet life" in ways that make the definition of "extremely improbable" match the 10^{-7} requirement, and then positing 100 critical systems (or failure modes) to arrive at 10^{-9} for each. Lloyd and Tye (1982), a respected study about aviation regulation, exemplifies this pattern. To arrive at the nine 9s, its authors posit a fleet of 200 aircraft, each flying 50,000 (5×10^4) hours before retirement, thereby creating a total fleet life of 10,000,000 (10^{-7}) hours. They then suppose 100 sources of catastrophic failure in each aircraft and implicitly equate that to

100 critical systems, each requiring a mean-time-to-failure of 10^9 hours (Lloyd and Tye 1982, 37). As if to underline the interpretive flexibilities of this calculation, however, the authors then—on the same page(!)—offer an incommensurable justification of the 10^{-7} requirement: positing a fleet size of 100 aircraft instead of 200 and a life of 90,000 hours instead of 50,000.[14]

Another formulation of the same calculation was offered to the author by an FAA chief scientific and technical advisor in 2005:

> The one-in-a-billion figure comes from estimating the fleet life for one aircraft model. . . . One operational year is close to 10^4 operational hours (it is around 8500 hours), assuming 24-hour operations (many aircraft operate two-thirds of the day, every day, almost year-round), 30-year service life gives you 3×10^5 for an aircraft-lifetime, and 3,000 aircraft in the fleet . . . you get 9×10^8, which is close enough to 10^9 operational hours per fleet-lifetime.[15]

Consider the variations between these various accounts. The 1982 formulation imagines a service life for each airframe of 50,000 hours (or 90,000) and a fleet size of 200 airplanes (or 100), while the 2005 formulation posits a service life of 300,000 ("3×10^5") hours and a fleet size of 3,000 airplanes. These are very meaningful differences. Where the first concludes that a specific airplane type will accrue 10 million flight-hours during its lifetime, the second concludes that it will accrue a billion. In doing so, moreover, the latter reaches 10^8 without accounting for the fact that airframes have multiple critical systems. Upon being alerted to this apparent oversight, the same correspondent reflected candidly on the changing nature of the industry. "The [nine 9s] figure was derived at a time at which no model [of airframe] was expected to be in service 30 years, or [to be manufactured] in multiples of a thousand. . . . Maybe we should go to 10^{10}."[16]

Maybe they should.

FIGURES OF MERIT?

In light of the variance and ambiguity evident in the definitions given here, it is difficult to ascribe much rigor or agency to the quantitative targets at the heart of type certification. The seemingly crucial standards against which regulators ostensibly measure the reliability of airframes are self-evidently social constructions: rhetorically compelling, perhaps, but not mathematically meaningful. We might think of the nine 9s as a "Goldilocks number": suitably impressive without being inconceivable. It is worth noting that the same "one in a billion" figure appears in other high-profile contexts where

the public demands extreme levels of certainty. It is the oft-stated reliability of DNA matching, for instance, and is similarly problematic in this context (Lynch and Cole 2002).

The less-than-rigorous foundations of certification's reliability requirements are not lost on experts themselves. One uncommonly forthright correspondent—an experienced engineer once employed by a leading manufacturer—described the nine 9s as "fatuous nonsense . . . designed to gull the public into believing that someone is actually producing a figure of merit."[17] The FAA presumably disagrees on the specific question of fatuousness, but it too sometimes will concede the broader point in its more private and esoteric discourse. In supplementary literature accompanying a redraft of the guidance that explains the numbers (FAA 2002b), for instance, it responds to what it describes as "misinterpretation, confusion, and controversy" regarding their application. In an almost parenthetical passage, tonally incongruous with the analysis that precedes it, the regulator clarifies that the numbers are intended to be "guidelines" only and stresses that they cannot replace or supersede engineering judgment when making airworthiness determinations (FAA 2002b, 25).

These kinds of caveats are especially common in the literature around certification's implementation (as opposed to its purpose). Here, it is relatively easy to find language that suggests a nuanced, practical understanding of technology assessment that is more congruent with finitist accounts of engineering than with its positivist public image. The FAA refers to key variables as "somewhat arbitrary" (e.g., FAA 2002b, 5; 2002c, 24), for example, and routinely flags the limitations of quantitative analyses (e.g., FAA 1982, 2; 2002b, 7, 25; 2002c, 23, 25). A 1983 investigation by the National Academy of Sciences (NAS) really grasps the nettle, bluntly stating that "the determination of design and engineering adequacy and product safety [in aviation] cannot be legislated in minute detail" (NAS 1980, 23). In such literature, numerical values are often "assigned" rather than "calculated," and assessments are more often required to be "convincing" rather than "correct."[18]

Such comments might look innocuous, but they have far-reaching implications for how we—publics, policymakers, and laypeople of all kinds—should construe aviation safety in general and the type-certification process in particular. The judgments they acknowledge speak to a tension between certification's messy realities on one hand, and its portrayal as a set of exacting requirements that airplane manufacturers must satisfy on the other. They

also point to a dilemma. Because if experts cannot precisely define the reliability required of a system, then determining whether the reliability of that system is satisfactory cannot be a wholly objective process. And if assessing the extreme reliability of a critical system cannot be a wholly objective process, then, as we have seen, it places seemingly impossible demands on the experts' subjective decision-making.

4.3 PRACTICAL DILEMMAS

THE CHALLENGE OF AMBIGUITY

Insofar as we construe type certification as a set of exacting requirements that airplane manufacturers must satisfy, then it is tempting to imagine that its flexibilities must diminish the difficulties of satisfying those requirements (as presumably, it is easier for manufacturers to claim compliance with ambiguous standards that bend accommodatingly to interpretation). And in many catastrophic technological contexts, this might be true. As we have seen, however, civil aviation experts are in a unique position stemming from the fact that their reliability claims will eventually be tested against real, statistically significant service. In this domain, if in few others, they cannot navigate their regulatory obligations by exploiting ambiguities in the rules, because their exploits would become obvious when planes failed.

This means that airframers really do have to design and build ultrareliable jetliners, irrespective of the certification bureaucracy. The nine 9s might be a social construction but it is far from being an overestimate. Jetliners undoubtedly *do* have critical systems on which their safety depends, and meeting modern expectations about aviation safety undoubtedly *does* hinge on those systems functioning for billions of hours between catastrophic failures. (Indeed, as we saw earlier in this discussion, parsing the FAA's numbers suggests that the failure performance required of such systems is, in practice, considerably higher than what certification formally requires.) In these circumstances, it is reasonable to imagine that manufacturers would welcome well-defined requirements, and any ambiguities in those requirements are better understood as a source of difficulty rather than opportunity.

Seen in this light, the vagaries of type certification's reliability targets exemplify both the finitist case against ultrahigh reliability assessment and the finitist paradox of aviation safety that challenges that case. The reliability of a jetliner hinges on experts' ability to know the behavior of its

most critical systems to an extraordinary degree of accuracy and confidence. Yet those experts must establish their knowledge with reference to fundamentally ambiguous metrics, based on dubious foundations that cannot be applied without subjective judgment and interpretation. Jetliners are demonstrably reliable, however (or have proved to be so in the past), so experts must be making those judgments and interpretations with a degree of accuracy that is incommensurate with the rigor of their rules and metrics. The question, therefore, is *how* they do this.

To begin to answer this question, we must look at the practical work of certification in action. Let us now turn to examining one of its most fundamental practices: testing.

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