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

5 WHEN THE CHICK HITS THE FAN: TESTING AND THE PROBLEM OF RELEVANCE

> Not everything that counts can be counted.
>
> —William Bruce Cameron

> Experience acquired with turbine engines has revealed that foreign object ingestion has, at times, resulted in safety hazards. Such hazards may be extreme and possibly catastrophic involving explosions, uncontrollable fires, engine disintegration, and lack of containment of broken blading. . . . While the magnitude of the overall hazards from foreign object ingestion are often dependent upon more than one factor, engine design appears to be the most important.
>
> —FAA Advisory Circular (FAA 1970)

5.1 BIRD-STRIKE TESTING

A PERSISTENT HAZARD

On October 4, 1960, Eastern Airlines Flight 375 struck a flock of starlings six seconds after taking off from Boston. The birds destroyed one of the Lockheed Electra's four engines and stalled two others. Starved of power at a vital moment, the injured aircraft yawed steeply to the left at 200 feet and plunged headlong into the shallow green water of Winthrop Bay. Sixty-two people lost their lives.

Flight 375 was far from the first airplane to be brought down by wildlife. Orville Wright, ever the pioneer, claimed the first documented "bird strike" in 1905, just two years after he made the first powered flight. (It could not have come as a huge surprise; his diaries record that he was chasing a flock

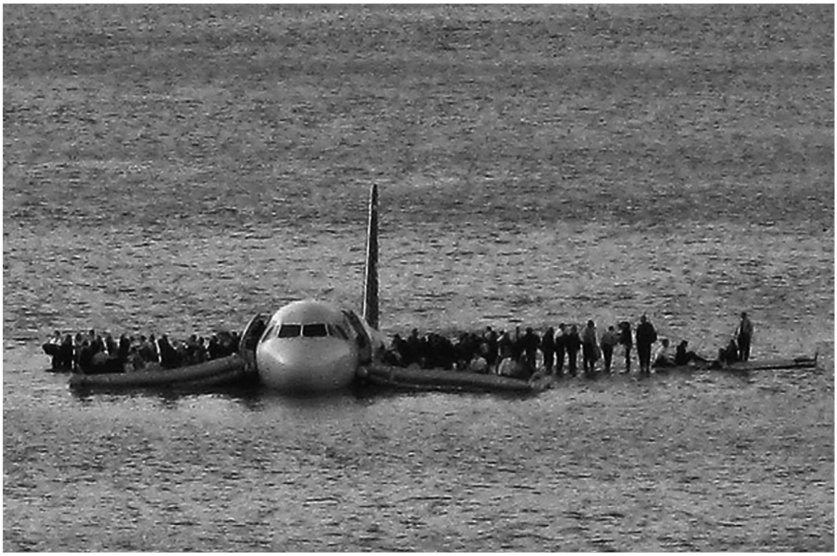

FIGURE 5.1
US Airways Flight 1549 landed in the Hudson River, January 15, 2009. *Source:* NTSB (2009, 5).

of birds around a cornfield at the time.) The first reported bird-strike fatality came seven years later, in 1912, when Cal Rodgers—a student of the Wrights and the first man to fly across the continental US—caught a seagull in his flight controls and plummeted into the Californian surf (Thorpe 2003).

Regrettably, moreover, Flight 375 was far from the last airplane to be brought down by wildlife. The years since the tragedy have failed to resolve what continues to be a problem. In January 2009, for instance, US Airways Flight 1549, an Airbus A-320, only narrowly avoided a similar catastrophe when a bird strike forced its pilot to ditch in the Hudson River next to midtown Manhattan (figure 5.1). Nobody died on that occasion, but aviation insiders were quietly astonished by the escape. Water landings are rarely so successful (Langewiesche 2009b). Globally, bird and other wildlife strikes killed more than 255 people and destroyed over 243 aircraft between 1988 and 2013 (FAA 2014, 1).

Birds are not the only wildlife threat to jetliners. Any creature that flys in the air or wanders onto runways is liable to be struck directly or sucked into an engine. The FAA Office of Airport Safety and Standards has published a table of animals struck by civil aircraft in the US between 1990 and 2013 (FAA 2014, 37–56). A veritable menagerie of American fauna, it lists forty-two species of terrestrial mammal and fifteen species of reptile, including 978 white-tailed deer, 443 coyotes, 243 striped skunks, 229 black-tailed jackrabbits, 19 alligators, 13 painted turtles, 5 moose, 3 horses, and 1 black bear. Elsewhere, there are even reports of at least one "fish strike"; albeit coincident with an osprey (Langewiesche 2009a). (Not astoundingly, there appears to be no established audit capacity for fish strikes.)

Such incidents can sometimes threaten the airplane, especially if they occur while it is taking off, but birds are by far the most prominent hazard. Bird strikes constitute over 97 percent of all reported wildlife strikes, with over 11,000 reported in 2013 alone (FAA 2014, x). The costs of these incidents are substantial. Birds destroyed over twenty US commercial aircraft between 1960 and 2003 (Bokulich 2003). In a 2014 report, the FAA estimated that they cost American carriers somewhere between $187 million and $937 million in damages annually, as well as between 117,740 and 588,699 hours of unintended downtime (FAA 2014, xi).[1]

Airports employ an innovative array of tactics to keep birds away from jetliners, but these are rarely more than marginally successful. Evolutionarily well designed for negotiating high fences but not for avoiding fast objects,[2] birds are often attracted to the open spaces around runways, where they then struggle to avoid the air traffic. Groundkeepers try to scare them off with a spectrum of creative measures—from plastic hawks and rubber snakes to lasers and pyrotechnics—but birds are not easily cowed, and many become inured to even the most vigorous attempts at intimidation. Scarecrows get claimed as nesting places; noise generators often double as popular perches.

Rather than bet too heavily on controlling the birds themselves, therefore, the onus is on the airplanes, which must be bird-resilient by design. This requirement falls most directly on the engines. Birds can damage any leading part of a jetliner, but the engines are most vulnerable for several reasons. They are more likely to be struck because they inhale so aggressively. The strike area that they present consists of precision machinery revolving at high speeds and pressures. (Protective grills over engine mouths are an engineering dead end. Any grill that was strong and dense enough to withstand birds at high speeds would occlude airflow to the turbines and pose its own risk of being smashed into the blades.) And their operation is safety crucial, especially during takeoff, when power demands are higher and pilots have less altitude with which to work.

For this reason, the FAA's engine certification requirements outline a range of mandatory bird-strike tests. These tests represent only a small fraction of the testing regimen involved in certifying an engine or airframe, but exploring them in depth helps illustrate a dilemma common to all tests, and in doing so illuminates the nature of reliability engineering itself.

BOOSTING ROOSTERS

In principle at least, the FAA's bird-strike tests for new engine designs are as straightforward as they are dramatic.

Engineers prepare by firmly mounting an engine onto an outdoor stand, where it is already an impressive spectacle. The largest engines—which weigh over eighteen tons and cost more than their weight in silver[3]—have mouths that yawn over thirteen feet (four meters) across. Inside are huge turbines with teeth of graphite and titanium, all balanced so delicately that a slight breeze will spin them on the tarmac like a child's paper windmill. Designed to operate at over 2,500° Fahrenheit [1,371° Celsius], well above the temperature at which most alloys melt, each blade in these turbines represents the very forefront of materials science.

With everything in place, the test begins. Engineers gradually apply power to the engine, and its turbines begin to turn. The giant fan blades spin faster and faster until the engine reaches maximum climbing speed, where it roars like a furious kraken and blows like an uncorked hurricane. At this point, the most powerful engines are gulping over 1,900 gallons of fuel an hour to produce up to 127,900 lbf (569 kN) of thrust, and the tips of their largest turbines are moving at close to the speed of sound. And then, into the maw of this howling monster—this technological jewel, the billion-dollar product of years of work by thousands of experts—they launch an unplucked, four-pound chicken.

Strictly speaking, the bird does not have to be a chicken; any bird of equivalent mass will do. And for the engine to meet its certification requirements, it must perform similar feats with other birds representing different weight categories. Along with the single large bird, represented by the chicken (4 pounds [1.81 kilograms], or 8 pounds [3.63 kilograms] for very large engines),[4] the regulations require a volley of eight medium-size birds (1.5 pounds [0.68

kilograms]), each fired in quick succession, and a further volley of sixteen small birds (8 ounces [85 grams]). (Many more will be launched in precertification tests [Bokulich 2003]). All are fired from a compressed-air cannon at takeoff speed, which is usually around 200 knots (230 miles per hour) (FAA 2000). (The US Air Force, for its part, tests its much-faster-moving airplanes with a sixty-foot cannon capable of firing a four-pound feathered bird, headfirst, at over 1,000 miles per hour. They are said to call it the "Rooster Booster.")

The pass/fail criteria are broadly the same in each test. If the turbine "bursts" (i.e., releases blade fragments through the engine cowl) when the chick hits the fan, or if it catches fire, or cannot be shut down properly, then the engine fails the test. The tests with smaller birds are more demanding. In these, engines also fail if their output drops by more than 25 percent or if they fail within five minutes of being struck (FAA 2000). (Although a drop below 75 percent is acceptable, so long as it doesn't exceed three seconds.)[5]

Once experts deem an engine to have passed the tests, the FAA holds that it has proved its ability to safely ingest all the birds that it will encounter in service. So far as the regulator is concerned, it has demonstrated its resilience to bird strikes, which cease to be a meaningful reliability concern.

Behind this technocratic surety, however, lie complex debates about what the tests actually reveal, as well as meaningful uncertainty about the confidence to be gleaned from passing them. Take, for instance, the pass/fail criteria outlined here. Although they may look relatively concise on the page, these criteria are less straightforward than they appear, as they are framed by multiple caveats, the application of which demand complex rulings. If two birds in a volley strike the same fan-blade, for instance, then the test can be deemed to be excessively challenging. Conversely, if the birds do not strike a predefined point of "maximum vulnerability" on the fan blades, then the test can be deemed insufficiently challenging. Such debates are worth examining, not because (or not *only* because) bird-strike tests are important in themselves, but because they offer generalizable insights into the wider epistemology of testing and its role in assessing the reliability of catastrophic technologies.

To understand the issues involved and their significance, it helps to begin by revisiting the finitist arguments about the limits of proof (see chapter 2) and elaborating on their application to technological testing.

5.2 REPRESENTATIVENESS

THE PROBLEM OF RELEVANCE

Bird-strike tests are what engineers sometimes call "proof tests": their purpose, as Sims (1999, 492) puts it, is "to test a complete technological system under conditions as close as possible to actual field conditions, to make a projection about whether it will work as it is supposed to." These projections are important; experts bet lives on their accuracy. But they are also problematic.

As we have seen, the relationship between tests and the phenomena they reproduce has long posed dilemmas for those who seek to establish the validity of engineering knowledge claims. This is because tests always involve some degree of extrapolation. Engineers must infer from them how the technology will perform in practice, and such inferences inevitably hinge on judgments about the tests' representativeness. This is to say that the validity of a test hinges on how accurately it reproduces the real-world conditions it ostensibly simulates (Pinch 1993; MacKenzie 1996b).

Experts cannot infer much about how a technology will behave in operation, in other words, unless they are confident that their tests *adequately represent* the actual conditions under which that technology will operate. To achieve this, they must either (1) design tests that simulate the real world accurately enough that the phenomena being investigated (such as an engine's bird-strike resilience) will present identically in both the test and the world; or (2) understand the meaningful differences between their test and the real world accurately enough to interpret their results in a useful way. Since even the most realistic tests always differ from the real world to some degree, however, satisfying either condition requires that experts determine which differences are meaningful and which trivial, but such determinations are inherently open-ended (see, e.g., MacKenzie 1989).

Reduced to its essence, the dilemma here is that there are a potentially infinite number of ways that a lab test might not match the real world, in all its nuanced and complex messiness, but testers can only hope to control—or even recognize—a finite number of variables (hence the term "finitism"). Scholars often refer to this dilemma as "the problem of relevance" (e.g., Collins 1982, 1988), but it goes by a variety of other names. Latour and Woolgar (1979) call it the problem of "correspondence"; and when Vincenti (1979) speaks of "the laws of similitude," Pinch (1993) of "projection," or Kuhn (1996 [1962]) of "similarity relationships," they are invoking the same underlying problem. Every FAA certification test necessarily grapples with this conundrum, but the debates around bird-strike testing amply serve to illustrate its nature and implications.

"CONSISTENT WITH SERVICE"

To understand the relevance debates around bird-strike tests, it helps to begin with the data on which the tests are premised. The FAA's current tests grew out of recommendations made in a 1976 National Transportation Safety Board (NTSB) report. The report, which examined a bird strike incident involving a DC-10 out of JFK Airport, suggested that all future engines be subject to "strike tests" using birds of sizes and numbers that were "consistent with the birds ingested during service experience of the engines" (FAA 2000). To satisfy this recommendation, the FAA set out to determine the nature of real-world bird encounters with a detailed study of real bird-ingestion incidents. It collected information on the types, sizes, and quantities of the birds involved, along with their effects on different engines.

These historical data became the basis for its bird-strike certification requirements. They were used to establish the conditions under which engines would be tested, the properties that tests would seek to examine, and the performance criteria that testers would be required to demonstrate. Over time, however, many expert observers have questioned their accuracy and appropriateness. There are too many such questions to fully document here, but a few prominent issues serve to illustrate the wider debate.

One significant cause for doubt arises from ambiguities in the collection process itself. At the most basic level, collecting bird-strike data involves counting and identifying birds killed by airplanes. This work is possible because the majority of bird strikes occur when aircraft are close to the ground (either taking off or landing),[6] meaning that investigators can gather remains from the runway. Not all strikes occur at ground level, however, and even a significant number of ground-level strikes go unrecognized or unrecorded (reporting is voluntary, and birds often pass through engines without anyone noticing), so the data are necessarily incomplete (Cleary, Dolbeer, and Wright 2006). (Over the period from 2004 to 2008, for instance, the FAA estimates that only about 39 percent of commercial aviation wildlife strikes were recorded, but that figure is up from about 20 percent in the 1990s [FAA 2014, 56]). This incompleteness is compounded by the fact that even recognized strikes present data-collection challenges. As might be imagined, the experience of being sucked through a roaring turbofan engine is traumatizing for birds. It turns them into a pulpy mess, known in the business as “snarge,” which spreads over a wide area. This makes individual birds difficult to find, count, and identify by species (and hence by vital characteristics such as mass) (MacKinnon, Sowden, and Dudley 2001).[7]

Further doubts arise from *ceteris paribus* concerns. This is just to say that some experts doubt whether historical bird-strike data, even if they were complete and accurate, would adequately represent future (or even present) conditions. Critics have questioned the FAA’s extrapolation from old to new engines, for instance, suggesting that data accumulated from incidents with early engines (and engine configurations) may not be applicable to newer ones (e.g., FAA 2014, 2). Others have pointed to changing bird populations. Ingestion standards are largely based on the assumption that bird numbers and distributions will remain constant over time, and yet the FAA itself has found that both are changing in ways that increase aviation hazards (FAA 1998; 2014, x). Bolstered by wildlife agencies, climate change, and shifts in farming techniques, many of the species that most threaten airplanes—such as gulls and pigeons, which feed on waste, and turkey vultures, which eat roadkill—have flourished in recent decades (Barcott 2009). Of the fourteen species of large-bodied birds in North America, for instance, thirteen experienced significant population increases over the thirty-five years from 1968 to 2003 (Dolbeer and Eschenfelder 2003). Perhaps most notably, the US-Canada goose population rose from 200,000 in 1980 to 3.8 million in 2013 (FWS 2013; Eschenfelder 2001). Gull and cormorant populations around the Great Lakes have similarly expanded, with the latter rising from just 6 nesting pairs in 1972 to over 230,000 by 2020 (Drier 2020). In line with these shifting demographics, the number of strikes reported by US airports increased significantly over the same period (FAA 2014; Barcott 2009).

Even if we assume that the data informing bird-strike tests are accurate, complete, and consistent, however, questions remain about the ability of testers to represent those data adequately. So a second set of debates surround the representativeness of the tests themselves.

RECREATING STRIKES

In many respects, the FAA is extremely cautious about the authenticity of its bird-strike tests. For example, it prefers using freshly killed birds to birds that were previously frozen because the former are considered more realistic. (Thawed birds can be dehydrated, and incompletely thawed birds can contain dense ice particles [FAA 1970, 8].)[8] This concern for authenticity is less fastidious in other respects, however, and critics have raised doubts about a wide range of test parameters. These concerns are too numerous to be outlined here in full, but highlighting a few paints a useful picture of the complications involved. To this end, therefore, let us briefly visit the debates around four distinct test parameters: size/mass, species, quantity, and speed.

SIZE/MASS As outlined earlier in this chapter, the certification standards require three sets of bird-strike tests, each with a different category of bird. These categories are labeled in reference to their size—large, medium, and small—but are defined by their mass. The "large" birds are either 4 pounds (1.8 kilograms) or 8 pounds (3.6 kilograms), depending on the size of the engine, the "medium" birds are 1.5 pounds (0.7 kilograms), and the "small" birds are 3 ounces (85 grams). The different categories are intended to represent the various species of bird that an engine might inadvertently swallow, but some experts question their correspondence with the birds that actually make it into engines.

Some birds that engines regularly encounter fit neatly into the FAA's categories. The small (3-ounce) bird is about the mass of the European starling, for example, which frequently falls prey to civil aircraft. And, although chickens themselves are rarely sucked into engines outside of laboratories, the large (4- or 8-pound) bird that they represent is about the same mass as various waterfowl, such as gulls, which are often found near runways. The medium (1.5 pounds) category, on the other hand, is less representative. Barred owls and red-shouldered hawks are approximately this mass, but aircraft rarely ingest either. Ducks, meanwhile, are a common engine ingestee, but they tend to range from 2 to 3 pounds (0.9 to 1.4 kilograms), thus falling between the test categories (Eschenfelder 2000).

Such discrepancies are more significant and contentious than might seem intuitive. Take, for example, a dispute that arose during the FAA's attempts to harmonize its standards with its European counterpart, the Joint Airworthiness Authority (JAA) (now the European Aviation Safety Authority [EASA]) (FAA 2000). The JAA had concerns that stemmed from tests that it had reviewed of an engine with fan blades made from a novel material. The tests had revealed something surprising. The new blades behaved similarly to the old ones when faced with the volley of "medium" birds and the single "large" bird. When faced with a bird of a size between "large" and "medium," however, they were found to be only "marginally equivalent" to previous designs, displaying an "inferior level of robustness" (FAA 2004). This led the European regulator to conclude that it could not straightforwardly infer how an engine would behave when struck by birds of different sizes from those on which it had been tested directly (FAA 2000).[9]

Another cause of contention lies in the fact that birds over 8 pounds are not considered in the tests at all, despite the wide variety of such birds that routinely encounter aircraft engines. Geese, storks, and swans all routinely reach greater weights—some exceeding 30 pounds—and, as noted previously, all three exist in large (and in some cases rapidly expanding) numbers. Flight 1549, which famously landed in the Hudson in 2009, was brought down by Canada geese, for example, as was a military Boeing 707 outside Anchorage in September 1995 (Marra et al. 2009).

The FAA contends that its tests adequately accommodate the risks posed by what it calls "very large" birds (i.e., those over 8 pounds) since the "large" bird test already assumes that the engine might be destroyed (defined as losing a single fan blade) and requires only that blade fragments be contained and the engine shut down safely. If the engine is destroyed safely by a large bird, the agency argues, it will also be destroyed by very large bird, and probably just as safely (BNME 1999; FAA 1998). However, this argument rests on two assumptions—that losing a single fan blade is the most catastrophic damage that a bird can cause; and that an engine is destroyed "safely" if the blade is contained by the engine housing—and critics point out that neither assumption is always borne out by experience. US Airways (1999), for example, has expressed doubt about both assumptions, citing a case where more than one blade was "liberated" after one of its aircraft ingested an Eider duck. The blades were contained as intended (which has not always been the case), but the strike seriously damaged the engine structure, almost breaking off the inlet cowl in a "potentially catastrophic" way.

SPECIES Other critics have concerns about the fact that the tests categorize birds by mass, but not by species. Aircraft in service ingest a wide variety of "avifauna"—from pelicans to parrots, 503 different bird species were involved in collisions with US aircraft between 1990 and 2013 (FAA 2014, vi; Langewiesche 2009a)—and it is easy to imagine that the strike effects of these birds might be determined by more than their mass alone. This fits with the discourse around bird strikes, which quite often highlights perceived distinctions between species. After an American Airlines flight ingested a cormorant in 2004, for instance, its spokesperson justified the large amount of damage by arguing that "a cormorant is chunkier, meatier and has more bones than a looser, watery bird [and] would have a harder time getting through the fan blades of the turbine" (Hilkevitch 2004).

The airline's assertion was certainly plausible. Different bird species have different shapes, consistencies, and densities, even when they have the same mass, and there are good reasons to believe that these variables significantly shape the birds' impact on engines (Budgey 2000). Take, for example, the aforementioned stipulation that a test bird (or birds) must strike the engine at its most vulnerable point for the test to be valid. To determine this point of vulnerability, engineers identify something called the "Critical Impact Parameter (CIP)." For most modern engines, this is usually the stress imparted to the leading edge of the fan blade (although other potential CIPs emphasize different engine elements, such as the blade root, and different variables, such as strain, deflection, or twist). Yet relatively small changes in the volume or density of the bird—such as those found among different species—affect the bird's "slice mass,"[10] which, in turn, can change the CIP and lead to a different test (FAA 2009, 3).

QUANTITY A further debate surrounds the numbers of birds that the tests use. The volleys of eight "medium" and sixteen "small" birds, each fired in quick succession, are intended to simulate a flock encounter. Their similarity to actual flocks is limited, however, and some experts worry this makes the tests unrepresentative.

The most prominent argument made in this context is that the numbers are too small to adequately represent the most challenging conditions an engine might face. Critics point out that various birds congregate in dense groups, causing engines to sometimes ingest more of them than the tests anticipate. This is especially true of birds in the "small" weight category, such as starlings, which can gather in prodigious numbers. An MD-80 is reported to have left the snarge of 430 dead starlings on a runway at Dallas, for instance, and a Boeing 757 in Cincinnati is similarly said to have left over 400 (Eschenfelder 2000, 4). The fact that tests only require a single large bird raises similar questions, as these also sometimes flock. Geese, swans, and storks all gather in large numbers, for instance, especially around migration time.[11] An engine that gulps one goose, therefore, is reasonably likely to strike others, making for a more demanding challenge than that for which it was tested. When Flight 1549 was pulled from the Hudson, for example, the remains of multiple geese were found in both its engines (NTSB 2009, 80).

The fact that bird numbers might be higher than the tests suppose has complex ramifications for various aspects of bird-strike certification. The FAA treats multiple strikes to the same blade as an anomaly, for example, and has justified this rule by arguing that the tests use considerably *more* birds than are likely to be encountered in an actual bird strike (FAA 2004).

SPEED Finally—for our purposes at least—there is a debate around the speed at which the birds should travel. As noted previously, certification guidance stipulates that they be launched at "takeoff speed," or about 200 knots, on the basis that the majority of bird strikes occur near ground level. Critics, however, point out that birds are still sometimes struck when aircraft are traveling higher, and therefore faster. (As of 2009, for example, the highest recorded bird strike was at 37,000 feet, with a griffon vulture off the coast of Africa [Croft 2009]). This has led some to suggest that 250 knots—the maximum airspeed allowed below 10,000 feet—would make a more challenging test (ALPA 1999). Others have pushed for even higher speeds.

Regulators have rebuffed these suggestions, claiming that bird-strike tests become less rather than more challenging when birds are traveling faster than 200 knots. The argument here again relates to the engine's Critical Impact Parameter. The FAA claims that the lower speed is more likely to "result in the highest bird slice mass absorbed by the blade at the worst impact angle, and therefore results in the highest blade stresses at the blade's critical location" (FAA 1998, 68641). Again, however, this is contested by many observers. The Airline Pilots Association (ALPA), for instance, questions whether the "slice mass" argument is a proven assumption. And the FAA itself has elsewhere concluded that strikes occurring above 500 feet—where airplanes are usually traveling faster than 200 knots—are more likely to cause damage than strikes at or below 500 feet (FAA 2014, xi).

MEANINGFUL DIFFERENCES

The misgivings outlined here are far from an exhaustive list of the concerns that experts have expressed about the representativeness of the FAA's bird-strike tests. Many other elements of the tests are contestable and contested. For example, there are debates about whether the explosions of air from the cannon affect the turbines and about whether the test engines should be attached to automatic recovery systems, as they would be under flight conditions (FAA 2001). Stepping back, we might also consider that the tests might be unrepresentative of the precision and care with which testers build and handle their test specimens: a dilemma common to many technological tests (Sims 1999, 501; MacKenzie 1989). The engines that experts use for certification testing are invariably new machines that have been pored over by scores of engineers, whereas those that encounter birds in real life will have seen some service and routine maintenance.

As should hopefully be clear by now, however, there are a potentially infinite number of ways in which bird-strike tests might imperfectly represent real bird strikes, and these potential discrepancies are important. The FAA has credible answers to many of the criticisms outlined here, but it has no *definitive* answers. The differences that critics have highlighted between the tests and real-world conditions might be irrelevant, but there are credible reasons to doubt this. Bird-strike experts can reasonably be confident that the colors of the birds and the day of the week are unlikely to affect their tests, but the issues outlined here represent meaningful uncertainties. Ultimately, there is no escaping the fact that the usefulness of bird-strike tests—what their results mean for an engine in service—hinges on a spectrum of contested relevance assumptions, concerning everything from the significance and accuracy of historical data, to the ways that engineers interpret that data and represent it in the laboratory.

The epistemological dilemmas that relevance assumptions pose for bird-strike tests do not stop here, however, for testers must also reckon with problems that arise from too *much* representativeness.

5.3 REPRODUCIBILITY

NARROWING OF VISION

Somewhat confoundingly, amid the concern surrounding the imperfect realism of bird-strike tests, there are critics who contend that the tests would be stronger if they were *less* realistic. Take, for instance, the following lament from a paper presented to the International Birdstrike Research Group:

> It has long been accepted that using real bird bodies in aircraft component testing is not ideal. The tests are not uniform. . . . Differences in bird body density between species and even between individuals of the same species may cause different and unpredictable effects upon impact, with consequent implications for testing standardizations throughout the world (Budgey 2000, 544).

Note how this argument invokes the variations among bird species outlined in this chapter, but it construes the problem that variation poses very differently: as something to be eliminated rather than faithfully reproduced.

This position is not as illogical as it might appear. All knowledge practices involve what Scott (1998, 11) describes as a "narrowing of vision," wherein experts seek to isolate key aspects of an otherwise "complex and unwieldy reality" by eliminating confounding variables. And STS scholars have long attested to the importance of such narrowing; arguing that it underpins every effort to classify, compare, and ultimately comprehend the world (e.g., Latour 1999, 24–79; Bowker and Star 2000). By this view, the usefulness of engineering tests stems, in part, from their ability to control the world so that engineers can isolate a single variable; which is to say their artificiality rather than their realism (e.g., Wynne 2003, 406; Henke 2000, 484; Porter 1994, 38). For instance, experts would struggle to assess the relative strengths of different metals if the tests of each metal were conducted at different temperatures or with samples of different purities.

Seen through this lens, it is important that bird-strike tests be reproducible. Partly because they would not be credible if they produced different results at different times: the same engine passing one day and failing the next. And partly because reproducibility offers insight: if every bird-strike test were unique, testers would struggle to compare the performances of different engines and would learn little about the factors that influence that performance. To be reproducible, however, the tests must be standardized; they must accurately represent each other, as well as the world itself. And achieving this means stripping away sources of variation—anything that "may cause different and unpredictable effects."

Hypothetically at least, it ought to be possible for engineers to maximize the reproducibility of most technological tests without compromising those tests' representativeness. They could do so simply by re-creating the variables that affect the phenomenon being tested, while controlling everything that does not: reducing the tests' realism (and thereby rendering them more legible and reproducible) without altering anything that matters to their outcome. This work would still force experts to grapple with the problem of relevance, as they would have to identify every significant variable, but it would be mean they could pursue reproducibility and representativeness at the same time.

When reliability is the specific variable being tested, as it is in bird-strike tests, however, then reproducibility is in direct conflict with representativeness. The reason for this harks back to reliability's uniqueness as an engineering variable. As outlined in chapter 2, reliability is inherently contextual: it is a measure of a system's failure performance in the real world and has no meaning independent of this context. This means that, unlike most engineering variables, it cannot be abstracted or isolated. In tests of reliability, therefore, reproducibility and representativeness inherently pull in opposite directions. The real world is inherently messy and complex, so any reduction of that variability—any narrowing of vision—is, almost by definition, a potential reduction in the test's relevance. In these circumstances, testers must choose whether to prioritize reproducibility or representativeness, knowing that each comes at a cost to the other.

In the specific context of bird-strike tests, these tensions are well illustrated by the debates that surround the use of artificial birds.

RUBBER CHICKENS

Engineers concerned with the reproducibility of bird-strike tests often advocate for the use of artificial birds. Usually made from gelatin and sometimes referred to as "cylinders of bird simulant material," artificial birds are the physical embodiment of an engineering ideal. This is to say that they are birds as seen through a lens that reveals only those variables deemed significant to ingestion tests. When Wile E. Coyote contemplates the Roadrunner in Warner Brothers' classic cartoons, he sees only dinner: roasted and aromatic. Bird-strike engineers, in much the same vein, would see a potentially hazardous mass with a specific volume and uniform density. Neither is interested in the subtleties.

Proponents of artificial birds often highlight advantages relating to their cost, convenience, and (usually obliquely) the intangible well-being and public relations convenience associated with firing fewer actual birds into engines.[12] Above all else, however, they emphasize the virtues of reproducibility. With their gelatinous simplicity, artificial birds promise to reduce the (literal and figurative) messiness of bird-strike tests: limiting the extraneous variables involved in each strike to make those that count more susceptible to measurement and calculation. They lack the feathers, bones, beaks, and wings of a real bird, and much else besides; but, in this view, such features are part of a complex and unwieldy reality that can be stripped away to make birds more legible.

The benefits of reducing the realism of bird-strike tests in this fashion undeniably come at a cost, however. Even proponents of artificial birds recognize that the projectiles do not accurately reproduce the complexities of a real collision, and this makes the tests a less faithful representation of the phenomenon they are intended to measure. This is to say that those proponents wish to remove certain variables, not because they are insignificant or extraneous, but because they are the source of significant variations in the data. Unsurprisingly, therefore, a lot of experts have misgivings about the representativeness of artificial birds, many of which map onto the debates about real birds outlined in this discussion, albeit usually in a more acute form. (As outlined here, for example, some experts consider "toughness" to be an important variable when modeling bird strikes, and yet artificial birds are not designed to have the same toughness as a real bird [e.g., Edge and Degrieck 1999]).

The debates around this are complicated by the fact that reproducibility can sometimes be more useful than representativeness: assessment bureaucracies often preferring a test that ignores significant variables to one that is difficult to repeat (Porter 1994). Indeed, I will argue in later chapters that type certification leans heavily on comparing new systems with the systems they are replacing, with greater reproducibility (facilitated by greater standardization) facilitates. (It being more useful for experts to explore the question "*how will this system perform in the world*" by approaching it through the question "*how does this system perform relative to its predecessor*," than by trying to explore it directly.) Perhaps counterintuitively, therefore, the costs of artificiality might be worth paying.

For the time being, however, let us simply understand the problem of reproducibility as a further epistemological complication: one more reason to doubt that bird-strike tests can offer the kind of certainty that experts impute to them.

5.4 AN ENDEMIC DILEMMA

"NOT ALL THE WAY"

Bird-strike tests are a modest fraction of certification's test regimen, but they illustrate a much more generalizable principle. Far from being exceptional, their ambiguities and uncertainties speak to an indelible gap between every certification test and the phenomena that those tests aspire to represent. (Later chapters will outline many similar debates and controversies around other systems and variables.) "Tests get engineers closer to the real world," as Pinch (1993, 26) puts it, "but not all the way." The relevance questions that permeate them are never solved. With a potentially infinite number of variables that might be significant in any test, experts cannot be certain they have identified them all or even that they have adequately represented the variables they identified.

Recognizing these uncertainties of representation, and the problem of relevance that they imply, is not the same as claiming that technological tests are irrelevant, valueless, or open to any interpretation.[13] Per chapter 2, however, it is important to understand that these uncertainties matter in contexts where experts are looking to establish extreme reliability, far more than they matter in other engineering contexts. Claims about reliability are essentially claims about *confidence*, and a small but crucial element of any claim about jetliner reliability is an extreme confidence in their engines' ability to ingest birds safely. That level of confidence is incompatible with the irreducible uncertainties of real bird-strike tests.

For engineers to properly grapple with ultrahigh reliabilities via tests they would need their tests to represent even the most esoteric variables: the kinds of obscure conditions and interactions that appear only once in every billion flights. Any expert claiming to have established a system's ultrahigh reliability in a test, therefore, would be asserting (with extreme certainty) that they had accurately reproduced all the most confounding and unusual conditions that system will face in the entirety of its messy operational life (those that occur only with "a specific phase of the Moon or something," as one correspondent put it). This is not realistic.

This is all to say that testing alone is a woefully insufficient explanation for the performance of modern aircraft, and for the FAA's ability to accurately predict that performance. A close examination of its real practices amply illustrates the finitist case against ultrahigh reliability management.

Still, however, the manifest insufficiency of testing in this regard does not diminish the empirical facts of aviation safety, which is as evident in the bird-strike data as anywhere. Catastrophic bird-strike incidents like Flight 1549 do occur, but there is no statistical basis to claim that engines are less bird resilient than experts claim. The actuarial data speak louder than any lab test: birds very rarely bring down commercial jetliners. If anything, the resilience of engines to birds appears to be increasing. The percentage of wildlife strikes that cause damage has declined from 20 percent in 1990 to 5 percent in 2013 (FAA 2014, xi). In this, as in countless other areas, therefore, aviation experts are transcending the limits of their tests.

The aviation paradox remains.

For a better answer to that paradox—a means by which experts could potentially be transcending the limitations of their tests—we might look to systems-level engineering, whereby engineers compensate for uncertainties in their designs (and their tests of those designs) by making them failure tolerant. It is to this that we now turn.

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