When the Chick Hits the Fan: Representativeness and Reproducibility in Technological Tests
Author(s): John Downer
Source: *Social Studies of Science*, Vol. 37, No. 1 (Feb., 2007), pp. 7-26
Published by: Sage Publications, Ltd.
Stable URL: http://www.jstor.org/stable/25474496
Accessed: 12-05-2015 03:07 UTC

S|S|S

ABSTRACT Before a new turbojet engine design is approved, the Federal Aviation Administration (FAA) must assure themselves that, among many other things, the engine can safely ingest birds. They do this by mandating a series of well-defined – if somewhat Pythonesque – 'birdstrike tests' through which the manufacturers can demonstrate the integrity of their engines. In principle, the tests are straightforward: engineers run an engine at high speed, launch birds into it, and watch to see if it explodes. In practice, the tests rest on a complex and contentious logic. In this paper I explore the debate that surrounds these tests, using it to illustrate the now-familiar idea that technological tests – like scientific experiments – unavoidably contain irreducible ambiguities that require judgments to bridge, and to show that these judgments can have real consequences. Having established this, I then explore how the FAA reconciles the unavoidable ambiguities with its need to determine, with a high degree of certainty, that the engines will be as safe as Congress requires. I argue that this reconciliation requires a careful balance between the opposing virtues of reproducibility and representativeness – and that this balance differs significantly from that in most scientific experiments, and from the common perception of what it ought *to be.*

Keywords aircraft engines, bird ingestion, civil aviation, technology regulation

When the Chick Hits the Fan:

Representativeness and Reproducibility in Technological Tests

John Downer

> 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. In addition, lesser but potentially severe hazards may involve airflow disruption with flameouts, lengthy or severe power losses, or momentary disruptions and possibly minor blade damage. 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. (Federal Aviation Administration, 1970)

In January of 1940, Henry Tizard, Churchill's influential science advisor, after witnessing a demonstration of Britain's first jet engine, wryly pronounced that 'a demonstration which does not break down in my presence

Social Studies of Science **37/1 (February 2007) 7–26**

ISSN 0306-3127 DOI: 10.1177/0306312706064235
www.sagepublications.com

is a production job' (Constant, 1980: 192). In the years since Tizard's indulgent benchmark, the job of demonstrating the virtues of a jet engine to the relevant authorities (in Britain and elsewhere) has become rather more demanding. Before the US Federal Aviation Administration (FAA) will approve a new engine design, the manufacturers must perform an extensive series of detailed and expensive tests to establish that it will operate reliably. This testing regimen – evaluating the engine under extreme temperatures (from –40 to >100°C) and all kinds of other environmental conditions – culminates in the dynamiting of a turbine blade while the engine runs at high speed, to see if it will tear through the engine housing.[1]

Edward Constant (1980: 20) argues that testing is the most significant application of the scientific method to technology. The exact relationship between scientific experiments and technological tests is complex and contentious, but the analogy has intuitive appeal. Both are conducted in circumscribed environments: experiments are employed to verify, confirm, or refute theories and to provide data on which theories can be built; while tests, we are led to believe, are an arena in which engineers can 'objectively and definitively' interrogate a technology, probe the usefulness of the models that informed its design, and reveal the 'truth' of its functioning (Constant, 1980: 21). The comparison is intuitive, unsurprisingly, considering how it is arguable that any clear distinction between the tests and experiments is tenuous at best – both being highly varied practices that are often difficult to categorically distinguish. Either way, the analogy is a productive one, and writers such as Trevor Pinch and Donald MacKenzie have already done much to establish the relevance of the sociology of scientific experimentation to technological testing: both can be treated as a process of argumentation and persuasion, and both have epistemological shortcomings (Pinch, 1993; MacKenzie, 1996). As Pinch writes: 'Tests get engineers closer to the real world but not all the way' (1993: 26).

Tests cannot carry engineers 'all the way to the real world' because, like experiments, they are laden with unavoidable ambiguities. The tests performed by the FAA are what engineers sometimes refer to as 'proof tests': their purpose being – in the words of Benjamin Sims (1999: 492) – '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.' Such 'projections' are always an exercise in inference, because we must infer from the test how the technology will perform in practice, and, as Pinch, Collins, MacKenzie and others observe, making such inferences inevitably involves a series of judgments.[2] These judgments, Shapin (1995: 261) argues, must be *credible:* for a test to be meaningful, its observers must be convinced that these judgments are sound and the test is genuinely representative.

For a test to be *representative* of the real world (and therefore *relevant*) it must either be sufficiently like the real world, such that the technology will behave the same way in both conditions, or any significant divergence from the real world must be understood well enough that the test results can be interpreted in a useful way (MacKenzie, 1996: 253). (Of course, one must also understand the 'real world' well enough to make these judgments.) A bird fired into an engine in a laboratory must be seen as a legitimate proxy for a genuine birdstrike. Since even the most 'realistic' tests will always differ in some respects from the 'real thing', engineers must determine which differences are 'significant' and which are trivial if they are to know that a test is relevant or representative. (In most engine tests, for instance, the speed at which the blades are turning is considered significant, whereas the color of the engine housing is not.) MacKenzie's (1989) study of missile testing illustrates the problems that arise when the similarity between test conditions and operating conditions is questioned. This is, of course, a principle long familiar to science studies; with Harry Collins – perhaps most famously – exploring and illustrating it in a series of papers (Collins, 1982, 1988; Collins & Pinch, 1998). It used to be known as 'the problem of relevance'. Latour & Woolgar (1979) call it the problem of 'correspondence', and when Vincenti (1979) speaks of 'the laws of similitude', Pinch (1993) of 'projection', or Kuhn (1977) of 'similarity relationships', they are invoking the same basic principle.

Relevance is not the only important dimension of a test or an experiment however. For either to be useful they must first be reproducible, and their results must be controlled and comparable. If a test or experiment is not designed to be representative of real-world conditions then it risks being dismissed as unrepresentative; but, as Henke (2000: 484) points out, credibility also depends on retaining some elements of the laboratory, such as controls and experimental methods. A balance must be struck between 'control' and 'authenticity' – between making tests accurate and making them convincing – it's part of what Henke (2000: 483) calls 'making a place for science'. In order to learn from tests it is often necessary to combine the results of one test with those of another, but, for this to be possible, they must be standardized and calibrated against each other. As Latour (1999: 36) writes, 'scientific practice entails the confrontation and negotiation of utter confusion'. This is to say that there is a degree to which experiments are useful because they are *not* like the real world: specific changes are made so that their circumstances are more controlled and their outcomes more legible. To achieve these goals we use our 'laws of similitude' to isolate a narrow range of variables, which we then try to replicate, measure, and vary with care. As James Scott explains:

> Certain forms of knowledge and control require a narrowing of vision. The great advantage of such tunnel vision is that it brings into sharp focus certain limited aspects of an otherwise far more complex and unwieldy reality. This very simplification, in turn, makes the phenomenon at the centre of the field of vision more legible and hence more susceptible to careful measurement and calculation. Combined with similar observations, an overall, aggregate, synoptic view of a selective reality is achieved, making possible a high degree of schematic knowledge, control, and manipulation. (Scott, 1998: 11)

Scott is discussing the knowledge required by government bureaucracies, but he might equally be talking about science experiments or testing jet engines.

If reproducibility and relevance are both key aspects of tests, however, they sometimes pull in different directions. It can be difficult to simplify a phenomenon without making it less relevant, and in this regard, I will argue, the analogy between scientific experiments and technological tests becomes strained. Both grapple with the same dilemma, but they do so under different circumstances. In what follows I will explore the dual demands of relevance and reproducibility as they manifest in the complex process of testing, and thereby certifying, jet engines. The specific tests I will focus on are those that the US FAA uses to examine an engine's ability to withstand the trauma of 'gulping' birds such as ducks and seagulls, which are often found near airports and are ingested with regrettable regularity by passing jetliners.

Bird Ingestion

In 1960, a large jetliner was leaving Boston's Logan International airport when it struck a flock of starlings. One of its four engines was destroyed, two others lost power, and unable to retain enough thrust to keep it airborne, the aircraft plunged into Winthrop Bay with the loss of 62 lives. In the years since 1960, 20 US commercial aircraft have been destroyed by birdstrikes (Bokulich, 2003). In 1999, the FAA announced that birdstrikes cost the US aviation industry US$327 million in damage and more than 500,000 hours of downtime each year (Air Line Pilots Association, 1999). The US Air Force – whose aircraft are at greater risk because they regularly fly at much lower altitudes – estimates that birdstrikes are responsible for killing two aircrew members every 3 to 5 years, downing two of their aircraft annually, and costing the service US$50 million to US$80 million each year (Feris, 2003).[3]

Not all of this damage involves birds being gulped by engines; some is the result of birds striking the fuselage of the aircraft, the leading edges of the wings, or the nose cone. Nevertheless, the engines are most at risk. Not only does the intense suction engines produce make them particularly vulnerable, but an engine's 'strike area' (where the birds hit) is filled with a mass of precision machinery revolving at high speeds and pressures.[4] Not all the damage involves birds, either. Experts around the world have voiced concern over the ingestion of a menagerie of creatures: iguanas, black-tailed prairie dogs, grasshoppers, mole-rats, warthogs, impalas, and much else besides. The engines literally suck them off the tarmac. Deer, in particular, frequently get onto runways in the western hemisphere (Feris, 2003). The following list – complied with taxonomic virtuosity by the National Wildlife Research Center – depicts the yearly average number of strikes to civil aircraft in the USA by wildlife other than birds:

- Ungulates – 46
- Carnivores – 21

- Bats – 6
- Turtles – 3
- Rabbits – 5
- Rodents – 5
- Alligators – 1
- Armadillos – 1
- Opossums – 2

Alligators notwithstanding, birds remain the most prominent threat to aircraft engines, and for obvious reasons. Efforts to keep them out of engines by keeping them away from airports have only been marginally successful, despite an innovate array of tactics ranging from plastic hawks and rubber snakes, to distress calls played over loudspeakers. Natural selection has yet to endow birds with an aversion to airports, and most become inured to even quite sophisticated attempts at intimidation: scarecrows are claimed as nesting places, and noise-generators double as popular perches. The onus, therefore, is on the engine: it must be bird-resilient, and all commercial aircraft engines must demonstrate their bird-resilience in a series of standardized tests overseen by the FAA.

The modern incarnations of these tests have their roots in a series of recommendations made in a 1976 report by the National Transportation Safety Board (NTSB) after an accident caused by bird-ingestion (Federal Register, 1977: 29688). The report suggested that all engines should be subject to a number of bird-ingestion tests, and that the 'numbers and the sizes' of the birds used should be 'consistent with the birds ingested during service experience of the engines' (Federal Aviation Administration, 2000). To meet this goal the FAA sponsored a detailed study of bird-ingestion incidents, collecting and comparing the types, sizes, and quantities of birds involved, and their effects on the performance of different engines. (Today, the worldwide bird-ingestion database includes data from 1970 through the present and encompasses more than 600 million flights.) This information was used to produce the bird-ingestion stipulations included in Federal Aviation Requirement Part 33 (FAR-33), which covers the design and construction standards for turbine aircraft engines. The stated objective of FAR-33's birdstrike requirements was to reduce the risk of hazard due to bird ingestion to at least 10 to the minus eighth power (10–8) per flight ('aircraft cycle').[5] That is equivalent to one 'hazardous' birdstrike for every 100 million takeoffs and landings. Because of their dramatic nature, the tests required by this standard have become the most widely known and oft-quoted aspect of the FAA airworthiness certification process.

The Test

In principle, the birdstrike tests stipulated by FAR-33 are quite straightforward – they simulate a birdstrike and measure the ability of the engine to cope with it – although a casual observer could be forgiven for thinking them a prank devised by a bored and mischievous (albeit rather resourceful) teenager. The procedure is simple. The engine is firmly mounted on an outdoor test stand where it is already an impressive spectacle. The most up-to-date engines – which consume more than 1900 US gallons (7200 litres) of fuel in 1 hour and produce up to 110,000 pounds of thrust (490 kN) – are more than 10 feet (3 m) in diameter, such that they yawn like open-mouthed whales. They drive air with huge rotating turbines made of long graphite and titanium fan-blades, balanced so delicately that a slight breeze will turn them as the engines lie idle.[6] When everyone is in place, the engine is gradually brought to maximum climbing speed, bellowing like an angry Kraken and blowing like an uncorked hurricane; the giant fan-blades spinning faster and faster until their tips are moving at close to the speed of sound. And then, into the mouth of this blender–furnace – this technological jewel, the product of millions of dollars of design effort and several years of work by hundreds of people – the engineers hold their breath, cross their fingers, and launch an unplucked 4 lb (1.8 kg) chicken.[7]

The chicken is fired from a compressed-air cannon at about 200 knots (370 km/h) – the approximate speed a plane would be traveling during takeoff and landing, when most birdstrikes occur.[8] This speed is contentious, however, with some critics arguing that birds are often struck when the aircraft is going faster (see, for instance, Air Line Pilots Association, 1999). The maximum allowed airspeed below 10,000 feet (3050 m) is 250 knots (460 km/h) and critics suggest this should be the speed for the test to represent the most challenging possible circumstances. The FAA contends that the test becomes *less* rather than more severe at speeds greater than 250 knots (460 km/h), so that the 200 knot (370 km/h) stipulation 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' (Federal Aviation Administration, 1998). This is also contentious. The Air Line Pilots Association is doubtful of the slice-mass argument, and questions whether it is a proven assumption. They also argue that the speed civil aircraft travel at low altitudes is rising beyond the 250 knot (460 km/h) limit.[9]

Several such bird-gulping feats are required for a large turbojet engine to meet its Airworthiness Certification Requirements. Along with the single 4 lb (1.8 kg) bird (recently raised to 8 lb [3.6 kg] for very large engines),[10] the FAA requires a volley of eight birds weighing 1.5 lb (0.7 kg), fired in quick succession, and a further volley of 16 smaller birds of 3 ounces (85 g) each – although many more will hit the fan in the course of designing an engine (Bokulich, 2003). In each test, the pass/fail criteria are broadly the same. If the turbine disintegrates or catches fire when the chick hits the fan, if it cannot be shut down afterwards, or if it releases fragments through the engine casing, the engine fails the test (Federal Aviation Administration, 2000). The tests with smaller birds are more demanding: engines also fail these tests if their output is reduced by more than 25%, or if they fail within 5 min of the ingestion (Federal Aviation Administration, 2000).[11]

These pass/fail criteria are concise on the page and seem clear, straightforward, and unambiguous in principle. In practice, however, determining whether an engine has passed or failed can involve a spectrum of equivocalities and complex judgments. Engines that look as if they failed may pass (if, for instance, two birds struck the same fan blade – an occurrence deemed 'unrepresentative' by the FAA), and engines that look as if they passed may still fail (if, for instance, the birds did not strike the right point in the engine).[12] Even after an engine has passed, there is still doubt about what 'passing' actually implies. The FAA holds that passing its bird ingestion tests means that an engine will safely and reliably ingest all the birds it is likely to encounter (to a given, and quantitatively defined, likelihood). As we shall see, however, some commentators disagree with the FAA's calculations and the implications they draw from them: they point to the kinds of ambiguities inherent in all technological testing, and argue that any claim to certainty masks an ocean of doubt.

Doubts

The first points of contention surround the information on which the tests are based. As we saw above, the tests are derived from data about previous bird ingestions by aircraft in service, taking into account the types of birds involved, the type of aircraft, the type of engine, and the outcome of the encounter. From this, the FAA derives information about what kinds of birds were most often ingested, the effects of different sizes of birds on different types of engines, and the likelihood that more than one engine would be struck at the same time. This is then used as the basis for the FAA's bird-ingestion standards, but some birdstrike experts have long contested the relevance and representativeness of these data.

The statistics are gathered by counting and identifying the birds killed when a birdstrike is recorded. This is possible because the vast majority of birdstrikes happen when the aircraft is close to the ground – either taking-off or landing – and investigators can collect the birds from the runway. It is still a difficult task, however. Birds often pass through engines without the pilot or anyone else noticing. When they *are* noticed, the birds (or bird) – traumatized by their misadventure – are often spread, in pieces, over a wide area, making them difficult to find, count, and identify by species (and hence by weight). Because of this, the data are imprecise and arguably unreliable (Budgey, 1999).

A more prevalent line of criticism lies in the familiar problem of induction: even if we assume the data are accurate, they still only offer an account of past experience and may not be relevant to future bird threats (Federal Aviation Administration, 2000). The FAA bird-ingestion standards are based on the assumption that 'the historical environment will not worsen', but there is little doubt that the number and distribution of certain birds are, in fact, changing significantly (Federal Aviation Administration, 1998). The number of Canada geese in the USA, for example, has quadrupled since 1987, and some estimate that it will soon double again

(Eschenfelder, 2001). The gull population around the Great Lakes, meanwhile, is suffering the effects of overcrowding, whilst the cormorants in the same area have 'risen from the grave': their numbers rising from six nesting pairs in 1972 to more than 100,000 in 2001 (Eschenfelder, 2001). Several industry experts have argued that the rising bird populations will lead to increasing engine failures. The FAA believes that these changes are impossible to predict, however, and that they have to make do with the data available (Reed & Martingdale, 1998). In a similar vein, other commentators have suggested that the data accumulated from incidents with old engines may not be applicable to new ones, especially those with unconventional designs. Although many of the basic principles of engine design have remained stable for some time, there have been a number of significant innovations, especially in the materials used. Different alloys have been replaced by carbon-fiber and alloy composites. As we shall see below, there is good evidence that induction from previous design is difficult in such circumstances.

Don't Count your Chickens

If we now assume the data that inform the tests are both accurate and relevant, questions still remain about the accuracy and relevance of the tests themselves. Take the birds used, for example. In some regards the FAA are quite cautious about the representativeness of the birds their tests require. Freshly killed birds are preferred to previously frozen ones, for instance, because they are considered more authentic. Frozen birds, if incompletely thawed, might contain dense ice particles that affect the test; even if thawed completely, they may have become dehydrated from the freezing (Federal Aviation Administration, 1970: 8). This concern for authenticity is less far-reaching when it comes to bird sizes. As we saw above, the FAA bird-ingestion standards require three sets of birdstrike tests: one with a single 'large' bird, a second with a volley of eight 'medium'-sized birds, and a third with a volley of 16 'small' birds. The three bird sizes are designed to be representative of the types of birds that an engine might inadvertently swallow, but the extent of their representativeness is questionable.

Some of the birds that engines regularly encounter do fit neatly into the FAA categories. The small (3 ounce [85 g]) bird is about the size of the European starling, which frequently falls prey to passing aircraft; and, although chickens themselves are rarely sucked into engines outside of the laboratory, the large bird (4–8 lb [1.8–3.6 kg]) that the chicken represents is about the same weight as various waterfowl, such as gulls, which are often found near runways. The medium (1.5 lb [0.7 kg]) category, on the other hand, seems less representative. Birds such as the Barred Owl or Red-shouldered Hawk are approximately this weight but aircraft rarely ingest either. Ducks, a common engine-ingestee, tend to range from 2 to 3 lb (0.9 to 1.4 kg), thus falling between the test categories (Eschenfelder, 2000).

The discrepancies between the sizes of the birds used in the tests and those encountered in the skies are more significant and contentious than it might first appear. Indeed, the exact sizes of the 'medium-sized' birds became a point of dispute in the FAA's attempts to harmonize standards with their European counterpart, the Joint Aviation Authority (or JAA) (Federal Aviation Administration, 2000).[13] The JAA had concerns that stemmed from their involvement in the testing of a large engine that had fan blades made from a new material and built to a novel design. What they found during these tests surprised them. The new fan blades behaved in a similar way to the old blades when faced with the usual volley of 'medium'-sized birds and the single 'large' bird, passing those tests easily. When faced with a bird of an intermediate size, somewhere between 'large' and 'medium', however, the fan blades were found to be only 'marginally equivalent' to previous designs, with an 'inferior level of robustness' (Federal Aviation Administration, 2004). This led the JAA and others to conclude they could not necessarily infer how an engine would behave when struck by birds of different sizes than those on which it had been tested (Federal Aviation Administration, 2000).

This conclusion is deeply antithetical to the underlying logic of the FAA ingestion standards, which silently presume a straightforward, linear relationship between most variables. In principle, therefore, this finding could also cast doubt on other aspects of the FAA certification regimen, such as the stipulation that 'a small bird ingestion test is not required if the prescribed number of medium birds pass into the engine rotor blades during the medium bird test', a rule that assumes medium-sized birds are always going to be more 'challenging' than smaller ones: a logical assumption, to be sure, but one that is rarely, if ever, examined directly (Federal Aviation Administration, 2000).

Even when the weights of the test birds do match those encountered in real flights, the fact that different species are often used – chickens instead of gulls, for example – is also sometimes seen as a problem. The tests stipulate the bird masses that should be used but not the species, yet different bird species have different shapes, volumes and densities, even when they are of the same mass. The aviation community regularly refers to differences between species when explaining birdstrikes. After an American Airlines flight ingested a cormorant, for example, American's spokesperson justified the large amount of damage by saying 'a cormorant is chunkier, meatier and has more bones than a looser, watery bird. … Once ingested by the engine, it would have a harder time getting through the fan blades of the turbine' (Hilkevitch, 2004). There is strong theory behind the view that differences in volume and density might be significant. Before the engineers can launch the birds, they must work out which engine part is most vulnerable, and how. They call this the 'critical impact parameter', or 'CIP', and it gives them an idea of where to aim the birds and what they should be measuring.[14] Relatively small changes in the volume or density of the bird – such as those found between different species – affect what engineers call the 'slice mass', which, in turn, can lead to a shift in the CIP and, consequently, to a different test.[15]

Another cause of contention lies in the fact that birds more than 8 lb (3.6 kg) are not considered in the tests at all, despite the wide variety of such birds threatening aircraft engines, including geese, storks, and swans, all of which routinely reach weights greater than 8 lb (3.6 kg), some reaching more than 30 lb (13.6 kg). These birds exist in large and growing numbers and are increasingly hazardous to aircraft. In September 1995, for example, a military Boeing 707 (E-3) was taking off from Anchorage when it struck a flock of Canada geese, ingesting them into two of its engines. Both engines were damaged, and the now uncontrollable aircraft plummeted into high terrain at the end of the runway, killing its 24 crew members (Eschenfelder, 2000).

The FAA is aware of the threat from very large birds, of course, but contends that its tests adequately allow for this possibility. It argues that, since the 4 lb (1.8 kg) bird test assumes that the engine might be destroyed anyway (by losing a fan blade) and only requires that the blade be contained by the engine housing and the engine be shut down safely, it doesn't matter how much bigger the birds get. 'If the engine is destroyed, safely, by a 1.8 kg [4 lb] bird', it writes, 'it will also be destroyed by a 6 or 8 lb [2.7 or 3.6 kg] bird and probably as safely' (Federal Aviation Administration, 1998; Bureau Navigabilité Des Moteurs et Equipements, 1999). This assumes, controversially, that losing a single fan blade is the most catastrophic engine damage that any bird can cause, an assumption that is not always borne out by experience.[16]

Another debate focuses on the number of birds used. The volleys of eight medium and 16 small birds are designed to simulate a flock encounter, but the extent to which they achieve this is debatable. Many of the birds that aircraft encounter, especially smaller birds such as starlings, tend to flock in very large groups. So when an aircraft encounters a flock, it can ingest a very large number of birds.[17] An MD-80 transport aircraft, for instance, left 430 dead starlings on the runway at Dallas. A Boeing 757 in Cincinnati is similarly said to have left more than 400. Even quite heavy birds have been ingested in large numbers; a USAir B-737 struck a flock of gulls near Daytona Beach and left more than 200 on the runway (Eschenfelder, 2000). In the tests, large birds are not part of a volley at all despite the large variety of birds 8 lb (3.6 kg) or larger, which travel in flocks. Geese, swans, and storks all flock – sometimes in large numbers – especially around migration time.[18] An aircraft that gulps one goose, therefore, is reasonably likely to strike others, and to be faced with a much more demanding test of its engines than is stipulated by the FAA.[19]

Birds of a Feather

Many other elements of the birdstrike tests are debatable (and debated), but it should be clear by this point that there are many ways in which firing birds into aircraft engines might not be exactly the same as in-flight bird ingestion. There are many other avenues through which the FAA's birdstrike standards can be, and are, contested. Already, however, we have seen myriad assumptions embedded in what is sometimes portrayed as a rigorous technological standard or a 'proof' of reliability. These include:

- changing bird populations are not significant;
- 200 knots (370 km/h) is the ideal ingestion speed;
- the number of birds used in the test is representative of the birds encountered in flight;
- types of birds, and their sizes, are representative;
- birdstrike data are accurate;
- volleys of birds are equivalent to flocks;
- there is a linear relationship between the sizes of the birds and damage they cause;
- the mass of the birds is more significant than their volume, shape, and density;
- losing a single fan blade is the most catastrophic consequence of a birdstrike;
- it is unlikely that two birds will strike the same fan blade.

And many others besides.[20] I will return to these ambiguities, and the potential shortcomings they represent to the FAA's ingestion testing. First a diversion. Because even though there are many arguments for why the tests might not be representative enough, there are also critics that take an opposite view and argue that the tests are *too* representative and need to be less so. The argument, in short, is that by moving towards greater representativeness, the FAA is liable to undermine other virtues such as reproducibility and control. This position is well illustrated by looking at the debate about artificial birds.

Rubber Chickens

When Wile E. Coyote contemplates the Roadrunner, in Warner Brothers' classic cartoons, he sees only dinner. Birdstrike engineers, in contrast, would see a potentially hazardous mass with a specific volume and density. Neither are much interested in subtleties. This is why artificial birds – sometimes referred to as 'cylinders of bird simulant material' – are only loosely based on the physical properties of real birds, having roughly the same size, mass, and overall density. Usually made from gelatin, they mimic real birds only in ways that they are deemed relevant to ingestion tests, their function being to simulate the bird impact rather than to copy the bird itself. They are birds as seen through a technocratic lens: reduced to their 'significant variables'. They might not have feathers, bones, beaks, or even wings, but they have the size, mass, and density of a bird, and so – in a system that only sees birds in terms of these three variables and is blind to anything else – they are identical to birds. Wings and feathers are part of the 'complex and unwieldy reality' that engineers ignore in order to make birds more 'legible'. The artificial bird, an oval ball of gelatin with a precise mass and density, is the physical embodiment of an engineering ideal.

All the major engine manufacturers have begun using artificial birds for parts of their pre-certification testing. The FAA even stipulates that artificial birds may be used in certification testing, as long as the inspector deems them to be an 'acceptable' equivalent, although, in practice, they are still controversial and certification tests are still done with real birds (Federal Aviation Administration, 2001: 3). Their advantages, stressed by their proponents, include cost, convenience, and the intangible moral well-being associated with a reduced need to launch actual birds into aircraft engines. More important, however, is the argument that artificial birds offer more standardization and reproducibility. Standardization and reproducibility have been difficult to achieve in birdstrike testing, and this has been a source of consternation. Take, for instance, this lament from a paper presented to the International Birdstrike Research Group:[21]

> 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)

An artificial bird mitigates many of these problems. Wings, feathers, and so on, are parts of a 'complex and unwieldy reality' that engineers can ignore in order to make birds more 'legible'. With its gelatinous simplicity the artificial bird is – to quote one manufacturer – 'more susceptible to careful measurement and calculation', and offers – to use Scott's words – a 'high degree of schematic knowledge, control, and manipulation' (Scott, 1998: 11).

It is easy to see why these traits would be valuable to the FAA. Reproducible results are clearly a useful property of any test that aims to be generalizable, and a standardized artificial bird would allow the FAA to be confident that it was conducting tests of equal stringency.[22] Standardized tests, moreover, would be useful and manipulable in a way that the tests otherwise would not be, because they would be comparable. In many ways, standardized tests and reproducible results are the first stage of any scientific or technical knowledge. As virtues, however, standardization and reproducibility come with shortcomings. Both are achieved by stripping the data of extraneous variables, and 'narrowing one's vision' only to what is significant. The danger, of course, is of stripping away something relevant to the phenomenon under investigation, whereupon the data – reproducible or not – become misleading.

It is widely recognized, even by their proponents, that artificial birds do not accurately reproduce the complexities of a collision with a real bird, and – importantly – that this affects the test. This is clear, for instance, in the quote above: 'Differences in bird body density between species and even between individuals of the same species may cause different and unpredictable effects upon impact'. 'Toughness' is one example: it is defined as the likelihood that the fan blades will 'slice' into a bird, and some see it as an important variable when modeling bird ingestion (Edge &

Degrieck, 1999). Most artificial birds are not designed to have the same 'toughness' as a real bird. Different, but related to this, is the issue of 'uniform density'. Although the density of artificial birds is based on the density of a real bird, it is the *average* density of a real bird. Artificial birds are normally a consistent density throughout, whereas real birds have bones, beaks, and talons: dense bits and less dense bits, sharp bits and gristly bits. As we saw above, in the discussion of 'chunky' 'meaty' cormorants, some engineers argue that this variable is significant. Indeed many of the dilemmas we have already seen apply doubly to artificial birds. Real birds have complexities that artificial birds ignore, and even the proponents of artificial birds concede that these affect the test.

Artificial bird proponents would use artificial birds to help transform the messiness of a bird in a jet engine into manipulable and comprehensible data. It is an appeal to the virtues of Scott's 'narrowing of vision': virtues such as intelligibility, utility, manipulability, and transferability, rather than realism or naturalness. The artificial birds allow us to create order from disorder; they limit the number of variables and reduce 'background noise' (Latour, 1999: 51). The justification for reducing a problem to a limited number of variables, however, is that only a few variables are relevant. But the proponents of artificial birds want to remove variables because they are the *source* of significant variations in the data. In cases such as this, the benefits of reproducibility come at an epistemic cost, it is impossible to make the test simpler without making it less realistic: we are trading representativeness for reproducibility.

This is Scott's critique of the 'narrowing of vision'. The justification for reducing a problem to a limited number of variables rests on the notion that only a limited number of variables is relevant. For example, engineers do not record the ambient light when testing an engine because they feel it doesn't affect the test. But when they suggest that certain variables should be removed – not just because they 'clutter' the data and hide significant variables – but because they are the *source* of significant variations in the data, the proponents of artificial birds are undermining this logic. If the particular oddities of each bird can cause 'different and unpredictable effects upon impact', then there is a strong case to be made that such oddities should be part of any test. Removing them is not the same as eliminating superfluous information such as the day of the week or the color of the engine.

As the many critics of artificial birds are keen to point out, therefore, swapping representativeness for reproducibility seems like a poor trade. They argue that the real issue in question is whether an engine can safely withstand a bird colliding with its innards at more than 200 miles per hour (320 km/h), and, as such, there isn't much to be gained from simplifying or standardizing the test. Standardized tests might allow the results to be compared and combined, but the FAA are not using the data to look for underlying patterns, or laws of nature, so concise manipulable data are not really a priority. Nor are they using the data to compare different engines: engines either pass the test or they don't – there are no gradations after that. So precisely standardized tests are a somewhat hollow virtue. The underlying argument is that standardization and reproducibility are being seen as intrinsic goods rather than as means to an end. As Brian Wynne (1988: 152) argues: 'Technology is not about universality as most philosophy of technology misleadingly suggests. It is about functioning in concrete, complex situations.'[23] What Wynne does not say, however, is that these situations often have to be considered and assessed in advance, and the assessments also constitute the reality in which a technology must exist.

One-in-a-Billion

Engine manufacturers are obliged to put an actual figure to the birdstrike reliability of their engines (one failure in every 100 million aircraft cycles [10–8]), and the FAA are required to verify it. The ambiguities in birdstrike tests make this incredibly difficult, because they make it hard to know if the tests accurately represent real bird encounters. Artificial birds, with even more question marks over their representativeness, would seem to make it even more difficult. Even without the artificial birds, however, it would be impossible to deduce an accurate, or even plausible, reliability prediction from the birdstrike tests. There are too many unknowns, too many compromises, to allow the FAA to deduce the 'one-in-a-billion' level of confidence they are obliged (by their mandate) to require from the tests. With so few tests and so many variables, the idea seems ridiculous. In theory this is a problem borne by the manufacturers, on whom the obligation rests to prove to the FAA that their engines will be reliable enough. In practice, however, the problem is shared: the FAA cannot simply reject each new design because of the epistemological challenges of proving its reliability, so both are in a bind.

The way they resolve this problem, in short, is by relating the tests to the information gathered from the service experience of the engine and of engines like it. When evaluating a new design, rather than approaching it ad novum, and asking, 'How reliable will this engine be?' they ask, 'How is this engine different from its predecessors?'. Then, they ask 'Will these differences make it more or less reliable than its predecessors?'. Now, because they have records of how often engines have failed in the past, they can – ceteris paribus – make a reasonable prediction about how it will perform in the future. Lloyd and Tye summarize this practice in their breakdown of reliability assessment for the British Civil Aviation Authority:

> In essence, [reliability assessments] require a background of actual service experience, relevant to the type of component, supported by detailed engineering evaluation and, where applicable, testing of the component. In short, where numerical estimates of probability would be meaningless, recourse must be had to traditional engineering processes to justify confidence that the exceedingly high reliability needed will be achieved. (Lloyd & Tye, 1982: 39)

Of course, this approach suffers from drawbacks of its own. There are disagreements over which data are relevant, and how they should be standardized. This approach has worked for civil aircraft, despite such drawbacks, because conditions have changed slowly and in predictable ways, and because engine designs are remarkably conservative: each new model evolves from those that preceded it. New engines are always modifications of old ones (as are airframes, and practically every other element in a new aircraft).[24] This is why Pratt and Whitney are keen to advertise their new engine as being 'based on the mature reliability of their previous model'. The huge reliance on service experience to validate ultra-reliable machines is why engineers are, generally speaking, both able and willing to vouch for the reliability of a new aircraft, but reluctant to vouch for that of a radically new system like the Space Shuttle.

This is important, because to use the tests in this way is, to some extent, to compare engines with other engines rather than with the 'real world'. It is still important that the tests be 'realistic', because the FAA are assuming that a better performance in the tests will translate into better performance in service. Change the tests too much, however, and it becomes increasingly difficult to tell if one engine has performed better or worse than its predecessors. If what the FAA really are measuring is the strength of an engine relative to previous designs, standardization might be more important than fidelity or representativeness. It would be counterproductive to change the FAA birdstrike tests to make them more representative of actual birdstrikes, if this meant they could no longer compare the performance of new engines against old ones.

In this light, the trade-off between fidelity and reproducibility advocated by the proponents of artificial birds takes on a different hue. Contrary to common perception, the tests are not isolated events with simple pass/fail criteria, but rather, they are iterated and comparative measures of generations of engines, used to compare one with another. If the tests are changed – even to make them more representative – it inevitably becomes more difficult to compare the performance of new engines with that of older ones. The data accumulated from the service experience of its predecessors then become less relevant, and the logic of the tests begins to unravel.

Because the optimal balance between reproducibility and representativeness inevitably plays out differently in these circumstances, the idea of tests as means of 'interrogating the world for deeper truths' begins to falter. There is a sense in which technological tests, when used in this fashion, are not so much a matter of *interrogating* the world as of *calibrating* it. They are yardsticks, significant because of their uniformity rather than because they reflect any inherent or 'natural' properties. They are a means of reducing ingestion tests to what Latour (1999) calls 'circulating references', standardized measures that can be systematized, compared, and analyzed (also see Weingart, 1991). As with the yardstick, what is important is not that the test represents a 'natural' measure but that it is constant, allowing us to measure one thing against another.

It follows that if we define experiments as attempts to probe the world in pursuit of deeper 'truths' (admittedly an idealized definition), attempts to construe birdstrike tests as experiments are potentially misleading. The world outside the laboratory is the arena in which the engines are genuinely interrogated for their birdstrike resilience, in a way more analogous with scientific experiments (as narrowly defined above). It is here that actual service data are accumulated that can then be used to inform the laboratory tests. Peter Weingart (1991: 6) has come to a similar conclusion about complex technologies in general: 'technical systems', he observes, 'turn into models for themselves: the observation of their functioning, and especially their malfunctioning, on a real scale is required as a basis for further technical developments and also for increasing their safety.' Insofar as we intend to deploy new but dangerous technologies with which we have no prior experience, the implications of this are complex and far-reaching; but therein lies a separate publication.

Notes

1. All the blade fragments must be contained by the engine casing for the engine to pass (Rozell, 1996).
2. Or what Bijker et al. (1987) call 'interpretive flexibility'.
3. Military aircraft can travel upwards of 500 miles per hour (800 km/h). At that speed the impact of a 4 lb (1.8 kg) bird lasts a mere 0.001 s, and can strike with a force exceeding 100,000 metric pounds (490 kN). Near the top of the Air Force's worry list is the unlikely turkey vulture. They are not stuck very often, but they are big, and make a salutary impression when they are hit. Although only involved in about 1% of incidents, they are reportedly the cause of about 40% of the total damage to aircraft (McKenna, 2003).
4. See also: *Birdstrike Discussion Forum,* 2004. A $US 2 billion Stealth Bomber was lost, and three crew members killed, on 28 September 1987 when the aircraft collided with a pelican over Colorado.
5. Flight cycles are used instead of flight hours because by far the most dangerous periods, as far as birdstrikes are concerned, are take-off and landing. For this reason the length of any given flight is largely irrelevant.
6. The blades, along with other engine parts, operate at 2500°F (1370°C), well above the temperature at which most alloys melt. They represent the very forefront of materials science (the metal elements are 'grown' as a single crystal) and manufacturing them is a delicate and esoteric art.
7. Sometimes a different bird of the same weight is used. I am told that pheasants are currently popular with one manufacturer. The birds are killed before they are used. Since no bird is going to survive being fired from a cannon, it would be pointless to struggle with loading a live one, as well as a little macabre.
8. One knot is a little faster than 1 mile per hour (1 knot = 1 nautical mile per hour = 6076 feet per hour, whereas 1 mile per hour = 5280 feet per hour), so 200 knots = 232 miles per hour (370 km/h). The US Air Force, for its part, has a 60 foot (18 m) cannon that will fire a 4 lb (1.8 kg) feathered bird, head first, at more than 1000 miles per hour (1600 km/h). They call it the 'rooster booster'.
9. The FAA Air Traffic Operations Office has run a test program in Houston that eliminates the 250 knot (460 km/h) speed limit below 10,000 feet (3050 m), and encourages 320–40 knot (600–30 km/h) climb speeds instead, as a 'capacity enhancement tool'. A Delta B-727 participating in this program struck some snow geese at 280 knots (520 km/h), sustaining severe damage (Air Line Pilots Association, 1999).
10. Only the largest engines are subjected to the largest birds because it is thought that smaller engines tend to ingest only portions of large birds, since the birds strike the engine structure and break-up before they hit the blades (Federal Aviation Administration, 1998).

11. A momentary drop below 75% is acceptable as long as it doesn't exceed 3 seconds.
12. The manufacturers argue that – based on probabilities appropriate to a flock encounter – multiple birds striking the same blade may be unrepresentative and overly conservative (Federal Aviation Administration, 1998).
13. This work commenced in 1989 and was managed by the Engine Harmonization Working Group. It includes representatives from the JAA, FAA, the Aerospace Industries Association (AIA) and the European Association of Aerospace Industries (AECMA).
14. For most modern turbofan engines, the CIP is the stress imparted to the leading edge of the fan blade. Other potential CIPs include the stress imparted to engine parts, such as the blade root, and different variables, such as 'strain', 'deflection', and 'twist'. The FAA offers these 'example considerations for determining the CIP' in one of its recent advisory circulars:

    > For Turbofan first stage fan blades, increasing the bird velocity or bird mass will alter the slice mass, and could shift the CIP from leading edge stress to some other highly stressed feature of the blade (for example, the blade root). For fan blades with part span shrouds, it may be blade deflection that produces shroud shingling and either thrust loss or a blade fracture that could be limiting. For unshrouded wide chord fan blades, it may be the trailing edge tip of the blade which experiences damage due to an impact induced shock wave traveling through the blade, or the twist of the blade in dovetail that allows it to impact the trailing blade resulting in blade damage. (Federal Aviation Administration, 2001: 4)

15. It is usually best if the blades can slice a bird into small bits before it passes through the engine. In the days before federal literature was rendered featureless and anodyne by oblique technical euphemisms, the FAA listed 'blades which effectively mince birds upon contact' among its 'desirable engine features' (Federal Aviation Administration, 1970: par. 3).
16. US Airways, for one, have expressed doubt about the fan-blade containment stipulation. They cite a case where several blades were 'liberated' after one of their aircraft ingested a 3.8 lb (1.7 kg) eider duck. Although the blades were contained, they seriously damaged the engine structure, almost breaking-off the 'inlet cowl' in a 'potentially catastrophic' way. This case casts doubt on both the 'single blade' assumption, and the assumption that an engine is safe so long as the blades are contained (US Airways, 1999).
17. Some of the logic behind the tests even rests on the assumption that the volleys use considerably *more* birds than are likely to be encountered in an actual birdstrike. This is the case in the reasoning behind the FAA's decision to treat multiple birdstrikes to the same blade as an anomaly that falls outside of test requirements (Federal Aviation Administration, 2004).
18. There are, of course, two migrations every year: the going out and the coming back. The US Air Force, never afraid of an authoritative euphemism, refers to them as 'waves of biomass'.
19. The FAA has, in fact, responded directly to the challenge that it requirements regarding bird mass and flock size are less severe than occur in nature. In its response it agreed that events can occur that are beyond the severity of the proposed requirements, but countered by arguing that it was not the intention to 'encompass the worst possible combination of all factors'; that these factors would be 'impossible to predict'; and would be 'beyond the capability of current engine technology'. It also pointed out that a number of new engine models have been designed and evaluated to standards, and had 'generally performed well in revenue service'. These arguments are no doubt valid, but their implications, if generalized, would seem to undermine much of the underlying logic of almost all of the certification process (Federal Aviation Administration, 2000).

20. These are not the only uncertainties, of course. Some critics worry that the explosion of air from the bird-launching cannon affects the engines because it is too close to the turbines; they question whether a bird fired from an air-cannon at 200 knots (370 km/h) is the same as a bird struck by an engine traveling at the same speed. There is also a debate about whether a proxy pilot should be allowed to increase the thrust on the test engine, or whether the engine should be attached to automatic systems, such as auto surge recovery and auto relight, as it would be under flight conditions. Another debate questions on whether testing should include only the moving engine, or other engine parts such as the exterior housing (see, for instance, Federal Aviation Administration, 2001). MacKenzie (1989) and Sims (1999: 501), moreover, both suggest that among most significant differences between test specimens and their real-world counterparts are simply the precision and care with which they are built and handled.
21. The International Birdstrike Research Group (IBRG) is a consortium of aerospace companies and other aviation organizations comprising BAE Systems, the UK Civil Aviation Authority, General Electric Aircraft Engines, The Gas Turbine Research Establishment, and India and Rolls-Royce Aerospace Group.
22. There is, however, little standardization of artificial birds throughout the world at present, although there are ongoing attempts to rectify this issue, not least by rival manufacturers of artificial birds (Budgey, 2000).
23. Benjamin Sims (1999: 502) describes a similar logic at work in a laboratory performing tests on building materials for earthquake resilience, arguing that engineers justified less realistic tests on the grounds that they 'enhanced the overall usefulness and applicability of test results'.
24. A good sense of the incremental and evolutionary nature of aircraft design can be got from reading the encyclopedic Loftin (1985).

References

Air Line Pilots Association (1999) 'Comments on Rules Docket (AGC-200)', Air Line Pilots Association International, Docket No. FAA-1998-4815 (52539) (March).

Bijker, Wiebe, Thomas P. Hughes & Trevor Pinch (eds) (1987) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology* (Cambridge, MA: MIT Press).

Bokulich, Frank (2003) 'Birdstrikes Remain a Concern for Pilots', in *Technology Update* (<www.sae.org/aeromag/techupdate_3-00/05.htm>, 11 May).

Budgey, Richard (1999) 'Three Dimensional Bird Flock Structure and Its Implications For Birdstrike Tolerance in Aircraft', paper given to International Bird Strike Committee (September) (IBSC24/WP 12).

Budgey, Richard (2000) 'The Development of a Substitute Artificial Bird by the International Birdstrike Research Group for Use in Aircraft Component Testing', *International Bird Strike Committee* IBSC25/WP-IE3 (Amsterdam, 17–21 April).

Bureau Navigabilité Des Moteurs et Equipements (1999) 'Commentaires NPRM 98-1 9', *Bureau Navigabilité Des Moteurs et Equipements*, Docket No FAA-98-2815-10 (53165) (31 March).

Collins, H.M. (1982) 'Tacit Knowledge and Scientific Networks', in B. Barnes & D. Edge (eds), *Science in Context: Readings in the Sociology of Science* (Milton Keynes, Bucks.: Open University Press): 44–64.

Collins, H.M. (1988) 'Public Experiments and Displays of Virtuosity: The Core Set Revisited', *Social Studies of Science* 18: 725–48.

Collins, Harry & Trevor Pinch (1998) *The Golem at Large: What You Should Know About Technology* (Cambridge: Cambridge University Press).

Constant, Edward (1980) *The Origins of the Turbojet Revolution* (Baltimore, MD: Johns Hopkins University Press).

Edge, C.E. & J. Degrieck (1999) 'Derivation of a Dummy Bird for Analysis and Test of Airframe Structures', in *Birdstrike 99 – Joint Birdstrike Committee USA and Canada Conference Proceedings* (Vancouver) (May).

Eschenfelder, Paul (2000) 'Jet Engine Certification Standards', paper delivered to International Bird Strike Committee (17–21 April) (IBSC25/WP-IE1).

Eschenfelder, Paul (2001) 'Wildlife Hazards to Aviation', paper given at ICAO/ACI Airports Conference (Miami, FL) (24 April).

Federal Aviation Administration (1970) 'Turbine Engine Foreign Object Ingestion and Rotor Blade Containment Type Certification Procedures', Federal Aviation Administration, Advisory Circular AC 33-1 B (22 April).

Federal Aviation Administration (1998) 'Airworthiness Standards; Bird Ingestion: Notice of Proposed Rulemaking', Federal Aviation Administration, (NPRM). CFR Parts 23, 25 and 33. Docket No. FM-1998-4815; Notice No. 98-18j RIN 21200AF34.

Federal Aviation Administration (2000) 'Airworthiness Standards; Bird Ingestion', Federal Aviation Administration 14 CFR Parts 23, 25 and 33; Docket No. FAA-1998-4815; Amendment No. 23-54, 25-100, and 33-20. RIN 2120-AF84 (Washington, DC, 5 September).

Federal Aviation Administration (2001) 'Bird Ingestion Certification Standards', FAA Advisory Circular AC 33.76-1 (19 January).

Federal Aviation Administration (2004) 'Response to Comments on NPA-E-20', Federal Aviation Administration (< www.jaa.nl/section1/crd/crd%20for%20npa%20e-20.doc >, 1 November).

Federal Register (1977) Vol. 42, No. 111 (June 9).

Feris, Melanie-Ann (2003) 'It's Murder on the Runways at SA's Airports', *Cape Argus* (1 September).

Henke, Christopher (2000) 'Making a Place for Science: The Field Trial', *Social Studies of Science* 30(4): 483–511.

Hilkevitch, Jon (2004) 'Goose Got the Blame but it was Rare Bird Plane Hit', *Chicago Tribune* (18 September).

Kuhn, Thomas S. (1977) 'Second Thoughts on Paradigms', in T.S. Kuhn, *The Essential Tension: Selected Studies in Scientific Tradition and Change* (Chicago, IL: University of Chicago Press): 293–319.

Latour, Bruno (1999) *Pandora's Hope: Essays on the Reality of Science Studies* (Cambridge, MA: Harvard University Press).

Latour, Bruno & Steve Woolgar (1979) *Laboratory Life: The Social Construction of Scientific Facts* (London: Sage Publications).

Lloyd, E. & W. Tye (1982) *Systematic Safety: Safety Assessment of Aircraft Systems* (London: Civil Aviation Authority).

Loftin, Laurence (1985) *Quest for Performance: The Evolution of Modern Aircraft* (Washington, DC: NASA Scientific and Technical Information Branch).

McKenna, Pat (2003) 'Birds Play Chicken with Air Force Jets' (< www.af.mil/news/airman/0296/duck.htm >, 11 May).

MacKenzie, Donald (1989) 'From Kwajalein to Armageddon? Testing and the Social Construction of Missile Accuracy', in David Gooding, Trevor Pinch & Simon Schaffer (eds), *The Uses of Experiment: Studies in the Natural Sciences* (Cambridge: Cambridge University Press): 409–35.

MacKenzie, Donald (1996) 'How Do We Know the Properties of Artifacts? Applying the Sociology of Knowledge to Technology', in R. Fox (ed.), *Technological Change* (Amsterdam: Harwood): 249–51.

Pinch, Trevor (1993) '"Testing – One, Two, Three ... Testing!": Toward a Sociology of Testing', *Science, Technology & Human Values* 18(1): 25–41.

Reed, Julian & I. Martingdale (1998) 'Birdstrike Statistics as a Design Tool', paper presented to International Bird Strike Committee.

Rozell, Ned (1996) 'The Boeing 777 Does More With Less', *Alaska Science Forum* (23 May) (< http://www.gi.alaska.edu/ScienceForum/ASF12/1286.html >).

Scott, James (1998) *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed* (New Haven, CT: Yale University Press).

Shapin, Steven (1995) 'Cordelia's Love: Credibility and the Social Studies of Science,' *Perspectives on Science* 3: 266–68.

Sims, Benjamin (1999) 'Concrete Practices: Testing in an Earthquake-Engineering Laboratory', *Social Studies of Science* 29(4): 483–518.
US Airways (1999) 'Comments on NPRM #FAA-1998–4815' Docket (53265) (18 Feb).
Vincenti, Walter (1979) 'The Air Propellor Tests of W.F. Durand and E.P. Lesley: A Case Study in Technological Methodology', *Technology and Culture* 20: 712–51.
Weingart, Peter (1991) 'Large Technical Systems, Real-life Experiments, and the Legitimation Trap of Technology Assessment: The Contribution of Science and Technology Studies to Constituting Risk Perception', in T.R. La Porte (ed.), *Social Responses to Large Technical Systems* (Dordrecht: D. Reidel): 5–17.
Wynne, Brian (1988) 'Unruly Technology: Practical Rules, Impractical Discourses and Public Understanding', *Social Studies of Science* 18: 147–67.

John Downer just graduated with a PhD from Comell's Science and Technology Studies Program. His thesis, written under the guidance of Trevor Pinch, was entitled 'The Burden of Proof: regulating ultra-high reliability in civil aviation' and won Comell's Guildford writing prize. This article is partially excerpted from it. As of this writing, he is enjoying a very rewarding but hopefully brief career as on over-educated unemployed person.

Address: 211 Stewart Avenue, Ithaca, NY 14850, USA; email: johndower@post.com.