Systems Thinking 1.0 and Systems Thinking 2.0: Complexity science and a new conception of "cause"

Sidney W. A. DEKKER[1]

ABSTRACT: *Our understanding and investigation of accidents in aviation is dominated by a mechanistic worldview that seeks to find and fix broken parts. Even though "systems thinking" has become quite fashionable over the past two decades, this still often reduces to finding broken parts further away in space and time from the accident. This is Systems Thinking 1.0. In contrast, complexity science, and its new conception of "cause," offers a route to Systems Thinking 2.0. In this, investigators and managers can be made aware of the consequences of path-dependence, open systems, the asymmetry between small inputs and large effects, and the unpredictability of efforts to control or regulate complexity. This paper uses a case study to compare Systems Thinking 1.0 and 2.0 and develop the latter.*
KEYWORDS: *Complexity, Systems Thinking, Accident Investigation, Aviation Alaska 261.*

RESUMO: *A nossa compreensão e investigação de acidentes aeronáuticos está dominada por uma visão mecanicista mundial a qual procura encontrar peças quebradas e reparálas. Apesar do "Systems Thinking" estar em voga nas últimas duas décadas, ele frequentemente ainda reduz a possibilidade de encontrar peças quebradas distantes no espaço e tempo do acidente. Esse é o Systems Thinking 1.0. Em contraste, a complexidade da ciência, e sua nova concepção de "causa", oferece um caminho para o Systems Thinking 2.0. Nele os investigadores e gerentes podem ficar a par das consequências da dependência, dos sistemas abertos, da assimetria entre pequenas contribuições e grandes efeitos, e da imprevisibilidade de esforços para controlar ou regular a complexidade. Este artigo usa um estudo de caso para comparar o Systems Thinking 1.0 e o 2.0 e desenvolver este último.*
PALAVRAS-CHAVE: *Complexidade, Systems Thinking, Investigação de Acidentes, Alaska Airlines voo 261.*

Introduction

Over the past two decades, accident investigations worldwide have attempted to

[1] PhD Professor, Co-Director Sir Samuel Griffith Institute Griffith University - Building M07 176 Mesiness Ridge Drive Mt. Gravatt, QL 412 Australia+61-(0)7-3735 5761 +61-(0)7-3735 6985 (fax) s.dekker@griffith.edu.au.

embrace a systems view of causation. In this view, accidents are not just the result of single broken components or badly performing humans, but rather a complex concatenation of a host of factors, most of which were around in the system for a long time before the accident. Despite these efforts, classical mechanics, as formulated by Newton and further developed by Laplace and others, retains a strong, if subtle, influence on how we think about accidents and their causation. This we can refer to as Systems Thinking 1.0. Our metaphors in safety, mostly driven by Cartesian-Newtonian thinking, can point us to the broken effects. They do a lot worse when it comes to understanding the socio-technical subtleties and normalities of daily organizational, bureaucratic life. The metaphors are mostly metaphors for resulting forms (e.g. broken links between organizational compartments, layers of defense with holes in them), not models that capture processes of evolution, formation, change. There are few workable models that capture the organic, jumbled, living interiors of the socio-technical organizations that govern aviation and other safety-critical activities.

Classical mechanics (and systems thinking 1.0) encourages a reductionist, mechanistic methodology and worldview. This means that the world gets seen as a series of machines with parts that interact and that can break. Even organizations get pictured as a collection of parts (e.g. layers of defense, stacked linearly). Finding out why an accident happens is often still a matter of finding out which parts were broken, even if we nowadays include a considerably larger number of parts (from wider and farther away). We now regularly look for sources of trouble in the organizational, administrative, and regulatory layers of the system, not just the operational or engineered sharp end. Indeed, this is what defines systems thinking 1.0 — finding more broken parts further away in time and space from the accident.

The continued pervasiveness of classical mechanics in these 1.0 attempts at systems thinking is not difficult to understand. The very legitimacy of accident investigation as a technical, scientific activity is at stake. Many, particularly in lay and engineering audiences, still equate "scientific thinking" with "Newtonian thinking." The mechanistic, Newtonian paradigm is compelling in its simplicity, coherence and apparent completeness. And it is consistent with people's intuition and common sense. The problem, of course, is that the language of classical mechanics is limiting at the same time that it is empowering. Even if it is dressed up as systems thinking, this language reveals some things, but hides even more. That is, the mechanistic view gives us strong hints of where to look for the causes of failure. But it also determines that we will not look in other places. It means that we will miss many other things, particularly the complex, non-linear interactions and emergent effects that classical mechanics (and Systems Thinking 1.0) has no language for at all. It is, however, the consensus of much accident research that the potential for disaster brews not just at the sharp, technical-operational end of a system. Rather, the potential for disaster develops as the

byproduct of normal social processes — organizational, administrative, managerial, political (Pidgeon & O'Leary, 2000; Turner, 1978; Vaughan, 1999, 2005). The nature of these processes may be hard to capture using Systems Thinking 1.0, and therefore this paper discusses a move to version 2.0, using a case study as illustration.

The case of a broken part

Here is an example. On January 31, 2000 Alaska Airlines flight 261, a McDonnell Douglas MD-83, crashed into the Pacific Ocean north of Anacapa Island, California. The 2 pilots, 3 cabin crewmembers, and 83 passengers on board were killed, and the airplane was destroyed by impact forces. Flight 261 was operating as a scheduled international passenger flight from Lic Gustavo Diaz Ordaz International Airport, Puerto Vallarta, Mexico, to Seattle-Tacoma International Airport, Seattle, Washington, with an intermediate stop planned at San Francisco International Airport. Visual meteorological conditions prevailed for the flight, which operated on an instrument flight rules flight plan.

As the investigation examined airplane parts from the sea floor and matched the wreckage with data traces from the cockpit voice recorder and the flight data recorder, the culprit became obvious: the jackscrew-nut assembly that holds the horizontal stabilizer had failed, rendering the aircraft uncontrollable. The broken part had been found. Like all conventional aircraft, the MD-80 has a horizontal stabilizer (or tailplane, or small wing) at the back that helps direct the lift created by wings. It is this little tailplane that controls the aircraft's nose attitude: without it, controlled flight is not possible. The tailplane of the MD-80 controls nose attitude in two ways. At the back end of it, there's a control surface (the elevator) that connects directly to the pilots' control yokes in the cockpit. Pull the yoke in the cockpit back, the elevator angles up and the airplane's nose moves up in return.

The story of Alaska 261, however, is about another part of the horizontal stabilizer. The whole stabilizer itself can angle up or down in order to trim the nose up or down. As fuel is used up during the flight, or wing flaps are extended and retracted, or as a catering trolley moves up and down the aisle inside the aircraft during flight, the flight characteristics of the airplane change (e.g. its center of gravity or the center of its lifting force shifts). To prevent the airplane from pitching down or up with changes in the center of gravity and center of lift, it needs to be able to trim. That's the role of the moving stabilizer. The stabilizer, that is, the entire horizontal tail, is hinged at the back, and the front end arcs up or down.

Here is how this is accomplished mechanically. Pushing the front end of the horizontal stabilizer up or down is done through a rotating jackscrew and a nut. The whole assembly works a bit like a carjack used to lift a vehicle, for example when changing a tire. You swivel, and the jackscrew rotates, pulling the so-called acme nuts inward and pushing the

car up. In the MD-80 trim system, the front part of the horizontal stabilizer is connected to a nut that drives up and down a vertical jackscrew. An electrical trim motor rotates the jackscrew, which in turn drives the nut up or down. The nut then pushes the whole horizontal tail up or down. On the 31st of January, twelve minutes after take-off, passing through 23,400 feet, the horizontal stabilizer had moved for the last time until the airplane's initial dive two hours and twenty minutes later.

Adequate lubrication is critical for the continued functioning of a jackscrew and nut assembly. Without enough grease, the constant grinding will wear out the thread on either the nut or the screw (in this case the screw is deliberately made of harder material, wearing the nut out first). The thread actually carries the entire load that is imposed on the vertical tail during flight. This is a load of around 5000 pounds, similar to the weight of a whole family van hanging by the thread of a jackscrew and nut assembly. Were the thread to wear out on an MD-80, the nut would fail to catch the threads of the jackscrew. Aerodynamic forces then push the horizontal tailplane (and the nut) to its stop way out of the normal range, rendering the aircraft uncontrollable in the pitch axis, which is essentially what happened to Alaska 261. Even the stop failed because of the pressure. A so-called torque tube runs through the jackscrew in order to provide redundancy (instead of having two jackscrews, like in the preceding DC-8 model). But even the torque tube failed in Alaska 261.

On the surface, the accident seemed to fit a simple category: mechanical failure as a result of poor maintenance. A single component failed because people did not maintain it well. But such accidents do not happen just because something suddenly breaks. There is supposed to be too much built-in protection against the effects of single failures. Other things have to fail too. More has to go wrong. And indeed, consistent with Systems Thinking 1.0, the investigation found more broken components in the system. Closest to the flight crew, it found that there was no suggestion in any checklist that the flight crew should divert to the nearest possible airport when they got the first indications of horizontal stabilizer trouble. In fact, they found that the use of the autopilot with a jammed stabilizer was "inappropriate" and that a lack of guidance on how to fly an airplane with a jammed stabilizer could lead crews to experimenting and improvising, possibly making the situation worse.

As for the lubrication of the jackscrew, the investigation determined that the access panel in the tail plane of this aircraft type was really too small to adequately perform the lubrication task. Also, there had been widespread deficiencies in Alaska Airlines' maintenance program, leading, for example to a lack of adequate technical data to demonstrate that extensions of the lubrication interval would not present a hazard. There was also a lack of a task-by-task engineering analysis and justification in the process by which manufacturers revise recommended maintenance task intervals and by which airlines establish and revise these intervals. Coupled to this, the investigation concluded that the end-play check

interval (which measures how much play or slack there is in the screw/nut assembly) was inadequate. A restraining fixture for end-play checks was used even though it did not meet aircraft manufacturer specifications. In addition, the so-called on-wing end-play check procedure was never validated and was known to have low reliability. There was no requirement to record or inform customers that allowed overhauled jackscrew assemblies back onto airplanes that had higher end-play than expected. Finally, the investigation noted shortcomings in regulatory oversight by the Federal Aviation Administration, and an aircraft design that did not account for the loss of the acme nut threads as a catastrophic single-point failure mode.

The Alaska investigation goes all the way up through the blunt organizational end and beyond, to the regulatory authority. But it still returns with broken components. This is the Systems Thinking 1.0 logic that has animated our understanding of failure for a long time now. Failure leads to failure. In order to explain a broken component (the jackscrew), we need to look for other broken components (the checklist, the access panel, the stress tests, the maintenance program, the company's oversight: all of them broken in their own way). In fact, any upbringing in Western science and engineering hardly allows people to think any other way.

Unanswered questions

But the idea that failure can be explained by looking for other failures leaves important questions unanswered. Questions, actually, that may stand between us and a firmer, further understanding of safety in aviation. Take, for example, the observation by the manufacturer that they knew of no reported cases of fatigue failures or fatigue damage on this torque tube design for their four basic models, the DC-9, MD-80, MD-90 and 717. No fatigue problems were reported in over 95 million accumulated flight hours, and in over 2,300 airplanes delivered (NTSB, 2002). That would suggest that nothing was broken — neither in the parts, nor in the organizational processes surrounding their maintenance and approval.

With the benefit of hindsight, it is not difficult to show that these parts were "broken." But why did none of these deficiencies strike anybody as deficiencies at the time? Or, if somebody did note them as deficiencies, then why was their voice apparently not persuasive? If things really were as bad as we can make them look in hindsight, then why was everybody, including the regulator — tasked with public money to protect safety — happy with what was going on? Happy enough, in any case, to not intervene? Clearly, people must have seen the norms that ruled their assessments and their decisions at the time as quite acceptable, otherwise pressure would have built for changing those norms (Vaughan, 1996).

These are issues at the heart of the future of aviation safety. If we don't understand why people see their decisions are normal, as safe, at the time, we will never be able to intervene meaningfully and prevent an accident such as this one. What we have to get at is this. Behind the broken part onboard Alaska 261 lay a vast landscape of things that weren't all that broken, or *not seen* as broken at the time. There had been organizational trade-offs and decisions that all seemed quite normal, there was deregulation and increasing competitive pressure that was normal because it operated on every company. There were changes in regulations and oversight regimes. Which there always are. There were underspecified technical procedures and continuous quality developments in aircraft upkeep, which are normal because procedures are always underspecified and changes in how to conduct maintenance are always ongoing — including the routine extension of maintenance intervals. There had been collective international shifts in aircraft maintenance practices, following the kind of cross-national public-private initiatives that mark a global industry. And all this interacted with what one company, and its regulator, saw as sensible and safe. So they kept on doing it. The decisions, trade-offs, preferences and priorities that people made, even if seemingly out of the ordinary and immoral *after* the failure, were once normal and common sense. Just like we believe that our decisions are today.

Introducing Systems Thinking 2.0

That the effects of apparently normal, everyday decisions can accumulate to become a disaster, is not something that relates directly to the decisions (or parts) themselves. In complex systems, there is no simple relationship between a broken part and a broken system. Systems Thinking 2.0 is about accidents that are more than the sum of the broken parts. This entirely reinvents the notion of cause. Rather than being "caused" by broken parts in a linear sense (like the breach of successive layers of broken defenses), accidents emerge from the multitude of relationships. Emergence is a conception of "cause" that is entirely different from Newtonian or mechanistic ideas. It is critical feature of complexity:

> Emergence is above all a product of coupled, context-dependent interaction. Technically these interactions, and the resulting system, are *non- linear.* The behavior of the overall system *cannot* be obtained by *summing* the behavior of its constituent parts. (pp. 121–2, original emphases). (Holland, 1998)

Any introduction of new rules, or new technology creates reverberations typical of complexity (Cilliers, 2002). New human roles emerge, and relationships between people and artifacts get transformed (Woods & Dekker, 2000). Interconnections between people and departments and artifacts proliferate. New kinds of human work are produced that call on new

sorts of expertise. While complexity theory does not provide the exact tools to solve such complex problems, it can provide rigorous accounts of the challenges that our safety work needs to meet (Cilliers, 2005). As Cilliers puts it (p. 258):

> […] because complex systems are open systems, we need to understand the system's complete environment before we can understand the system, and, of course, the environment is complex in itself. There is no human way of doing this. The knowledge we have of complex systems is based on the models we make of these systems, but in order to function as models—and not merely as a repetition of the system—they have to reduce the complexity of the system. This means that some aspects of the system are always left out of consideration. The problem is confounded by the fact that that which is left out, interacts with the rest of the system in a non-linear way and we can therefore not predict what the effects of our reduction of the complexity will be, especially not as the system and its environment develops and transforms in time.

Accident investigation has largely been constructed as a technical problem-solving activity, which leaves out deeper social or complex questions about the nature of work and organizations. As Wilkin (2009, p. 4) sums up (and consistent with the mechanistic worldview), such an "approach assumes an atomistic social world composed of variables… that [it] is a closed system of study where the individual parts (variables) can be separated, measured and controlled." Investigations then assume that they can meaningfully work with the units of a closed system that it has identified (Roesler, Feil, Puskeiller, Woods, & Tinapple, 2001; Wilkin, 2009; Woods & Dekker, 2000) but this is inconsistent with complexity. Here are some of the salient features of Systems Thinking 2.0 that need to be taken into account instead:

- Complex systems are open systems — open to influences from the environment in which they operate and influencing that environment in return.
- In a complex system, each component is ignorant of the behavior of the system as a whole, and doesn't know the full effects of its actions either. Components respond locally to information presented by them there and then. Complexity arises from the huge, multiplied webs of relationships and interactions that result from these local actions.
- Complexity is a feature of the system, not of components inside of it. The knowledge of each component is limited and local, and there is no component that possesses enough capacity to represent the complexity of the entire system in that component itself.
- Complex systems operate under conditions far from equilibrium. Inputs need to be made the whole time by its components in order to keep it functioning. Without that constant flow of actions, of inputs, it cannot survive in a changing environment.

- Complex systems have a history, a path-dependence. Their past is co-responsible for their present behavior, and descriptions of complexity have to take history into account.
- Interactions in complex systems are non-linear. That means that there is an asymmetry between for example input and output, and that small events can produce large results, in part because of the existence of feedback loops.

What do these features mean for our analysis of an accident like Alaska 261? The remainder of the paper is dedicated to a revisionist account along the lines of Systems Thinking 2.0. It covers the consequences of path-dependence, open systems, the asymmetry between small inputs and large effects, and the unpredictability of efforts to control or regulate complexity.

History matters: Path-dependence

When it first launched the aircraft in the mid-1960s, Douglas recommended that operators lubricate the trim jackscrew assembly every 300 to 350 flight hours. For typical commercial usage, that could mean grounding the airplane for such maintenance every few weeks. Immediately, the sociotechnical, organizational systems surrounding the operation of the technology began to adapt. Through a variety of changes and developments in maintenance guidance for the DC-9/MD-80 series aircraft, the lubrication interval was extended. These extensions were hardly the product of manufacturer recommendations alone, if at all. A much more complex and constantly evolving web of committees with representatives from regulators, manufacturers, subcontractors, and operators was at the heart of a fragmented, discontinuous development of maintenance standards, documents, and specifications. Rationality for maintenance-interval decisions was produced relatively locally, relying on incomplete, emerging information about what proved to be, for all its deceiving simplicity, unruly technology. Each decision was locally rational, making sense for decision makers in their time and place. But these local decisions eventually had an impact on the whole system.

Starting from a lubrication interval of 300 hours, the interval at the time of the Alaska 261 accident had moved up to 2,550 hours, almost an order of magnitude. This distance was not bridged in one leap but incremental: step by step, decision by decision that weaved through the aviation system from different angles and directions, alternating cross-industry initiatives with airline-driven ones, but all interacting to push the system into one direction: longer intervals.

In 1985, as an accompaniment of deregulation in the airline industry, jackscrew lubrication was to be accomplished every 700 hours, at every other so-called maintenance B check (which occurs every 350 flight hours). In 1987, the B-check interval itself was increased to 500 flight hours for the entire industry, pushing lubrication intervals to 1,000 hours. In 1988, B checks were eliminated altogether, and tasks to be accomplished were redistributed over A and C checks. The jackscrew assembly lubrication was to be done each eighth 125-hour A check: still every 1,000 flight hours. But in 1991, A-check intervals were extended across the entire industry to 150 flight hours, leaving a lubrication every 1200 hours. Three years later, the A-check interval was extended again, this time to 200 hours. Lubrication would now happen every 1,600 flight hours. In 1996, Alaska airlines removed the jackscrew-assembly lubrication task from the A check and moved instead to a so-called task card that specified lubrication every 8 months. There was no longer an accompanying flight-hour limit. For Alaska Airlines, eight months translated to about 2,550 flight hours.

The jackscrew recovered from the ocean floor, however, revealed no evidence that there had been adequate lubrication at the previous interval at all. It might have been more than 5,000 hours since it had last received a coat of fresh grease. Through this drift, the effect of small things could suddenly explode into something huge. Miss one lubrication if you service the part every 350 hours, and you have no worries. Miss one when you lubricate every 2,550 hours, and you could get in deep trouble. This is where small changes can lead to big events. And where the way the system is organized (e.g. around set lubrication intervals rather than continuous checks and lubrication-as-necessary) can make it highly dependent on such small changes.

Open systems

No organization is a closed system or operates in one. Instead, organizations operate and must try to survive in an environment that has real constraints (Rasmussen, 1997). Before the mid-1970s, the airline industry in most countries was essentially cartelized. Regulators assigned routes to airlines, and controlled the prices on these routes, resulting in strict limits on competition. Deregulation from the late 1970's onward changed all that. Market forces could now determine who could enter the industry, who could fly which routes and for what prices. Interestingly, deregulation became an asynchronous process in most countries: even though airlines were now allowed to compete under market forces, the infrastructure that they used remained in the hands of governments for a long time (airports, airways and the air traffic control system to run it), and even transnational arrangements often constrained heavily who could fly from which country to which. In the wake of deregulation, however,

the airline industry expanded its employment by 32 percent, and passenger travel increased by 55 percent. The real cost of travel dropped by about 17 percent in the first decade after deregulation alone, and dropped even further in the ensuing decades (Poole & Butler, 1999).

Deregulation cannot in itself be construed as a safety threat, of course. Heavily regulated industries do not necessarily have a better safety record (if anything, it may encourage collusion between regulator and industry, particularly if part of the regulator's role is to encourage business development), and the period after deregulation actually saw a steady increase in airline safety (Poole & Butler, 1999). But changing the rules of the game does change what goes on inside of a complex system. Complexity theory predicts that changing the number of agents will change the dynamics of any complex system; it will affect the speed at which feedback about agents' actions travels and the patterns along which it reverberates. It might even change the way in which success is defined and assured.

With respect to the Alaska 261 accident, for example, a new regulatory inspection program, called the Air Transportation Oversight System (ATOS), was put into use in 1998 (two years prior to the accident). It drastically reduced the amount of time inspectors had for actual surveillance activities. A 1999 memo by a regulator field-office supervisor in Seattle suggested how this squeezed their oversight work into a corner where the safety and workload boundaries met:

> We are not able to properly meet the workload demands. Alaska Airlines has expressed continued concern over our inability to serve it in a timely manner. Some program approvals have been delayed or accomplished in a rushed manner at the "eleventh hour", and we anticipate this problem will intensify with time. Also, many enforcement investigations … have been delayed as a result of resource shortages. [If the regulator] continues to operate with the existing limited number of airworthiness inspectors … diminished surveillance is imminent and the risk of incidents or accidents at Alaska Airlines is heightened (NTSB, 2002).

Adapting to resource pressure, approvals were delayed or rushed, surveillance was reduced. Yet doing business under pressures of resource scarcity is normal: Scarcity and competition are part and parcel even of doing inspection work. Few regulators anywhere will ever claim that they have adequate time and personnel resources to carry out their mandates. Yet the fact that resource pressure is normal does not mean that it has no consequences. The pressure finds a way out. Supervisors write memos, for example. Battles over resources are fought. Trade-offs are made. The pressure expresses itself in the common organizational, political wrangles over resources and primacy, in managerial preferences for certain activities and investments over others, and in almost all engineering and operational trade-offs between strength and cost, between efficiency and diligence (Hollnagel, 2009). In fact, working successfully under pressures and resource constraints is a source of professional pride. Being

able to create a program that allows better inspections with fewer inspectors may win a civil servant a promotion, while the negative side effects of the program are felt in some far-away field office.

Feedback imbalances: Learning the wrong thing

In making these trade-offs, however, there is a feedback imbalance. Information on whether a decision is cost-effective or efficient can be relatively easy to get. An early arrival time is measurable and has immediate, tangible benefits. How much is or was borrowed from safety in order to achieve that goal, however, is much more difficult to quantify and compare. If it was followed by a safe landing, apparently it must have been a safe decision. Extending a lubrication interval similarly saves immediately measurable time and money, while borrowing from the future of an apparently problem-free jackscrew assembly. Remember, the manufacturer knew of no reported fatigue problems or failures in 95 million flight hours accumulated by 2,300 airplanes.

Evidence from a feedback imbalance suggests strongly that the system can operate equally safely, yet more efficiently. From the outside, such fine-tuning constitutes incremental experimentation in uncontrolled settings (Starbuck & Milliken, 1988; Weingart, 1991; Wynne, 1988). On the inside, incremental nonconformity is an adaptive response to scarce resources and production goals. This means that departures from the norm become the norm. Seen from the inside of people's own work, deviations become compliant behavior. They are compliant with the emerging, local ways to accommodate multiple goals important to the organization (maximizing capacity utilization but doing so safely; meeting technical or clinical requirements, but also deadlines).

The brewing of an accident, however, hides somewhere in the conflicts that get sorted out in these trade-offs (Pidgeon & O'Leary, 2000; Rasmussen, 1997; Vaughan, 1999). This tension can lead to a slow, steady disengagement of practice from earlier established norms or design constraints (Leveson, 2011), and a redefinition of what is “normal” or “safe” or “acceptable” (Vaughan, 1996).

Small change in initial conditions, large effects

In complexity and systems thinking, the link between small changes in initial conditions and eventual large effects, is called a sensitive dependency on initial conditions (or butterfly effect). The original DC-9 was certified in 1965 under Civil Aeronautics Regulations (CAR) 4b from three years prior (1962). The MD-80 was certified in 1980. More recent models of the DC-9, MD-80, MD-90, and Boeing 717 were certified under 14 CFR Part 25

and applicable amendments. However, systems that were similar to or that did not change significantly from the earlier DC-9 models, such as the longitudinal trim control system, were not required to be recertified. CAR 4b remained the certification basis for those parts of the MD-80, MD-90, and 717. This meant that even the most modern of the lineage, the Boeing 717, today has parts in it that are certified in 1965 under rules from 1962, and never needed to be re-certified. Just a time check: these rules were laid down a year before Kennedy was assassinated. When the first DC-9 was certified, we still had four years to go before landing on the moon. These were slide rule times. The effects of, and sensitive dependency upon, such initial conditions are with us in many airplanes flying around today.

Understanding how a system may sensitively depend on initial conditions is something that certification processes are supposed to be able to do. But certification does not typically take lifetime wear of parts into account when judging (in this case) an aircraft airworthy, even if such wear will render an aircraft, like Alaska 261, quite unworthy of flying. Systemic adaptation or wear is not a criterion in certification decisions, nor is there a requirement to put in place an organization to prevent or cover for anticipated wear rates or pragmatic adaptation, or fine-tuning.

As a certification engineer from the regulator testified, "Wear is not considered as a mode of failure for either a system safety analysis or for structural considerations" (NTSB, 2002, p. 24). Because how do you take wear into account? How can you even predict with any accuracy how much wear will occur? McDonnell-Douglas surely had it wrong when it anticipated wear rates on the trim jackscrew assembly of its DC-9. Originally, the assembly was designed for a service life of 30,000 flight hours without any periodic inspections for wear. But within a year, excessive wear had been discovered nonetheless, prompting a reconsideration.

The problem of certifying a system as safe can become even more complicated if the system to be certified is sociotechnical and thereby even less calculable. What does wear mean when the system is sociotechnical rather than consisting of pieces of hardware? In both cases, safety certification should be a lifetime effort, not a still assessment of decomposed system status at the dawn of a nascent technology. Safety certification should be sensitive to the co-evolution of technology and its use, its adaptation. Using the growing knowledge base on technology and organizational failure, safety certification could aim for a better understanding of the ecology in which technology is released — the pressures, resource constraints, uncertainties, emerging uses, fine-tuning, and indeed lifetime wear.

The jackscrew in the DC-9 trim assembly had been classified as a "structure" in the 1960s, leading to different certification requirements from when it would have been seen as a system. The same piece of hardware, in other words, could be looked at as two entirely different things: a system, or a structure. In being judged a structure, it did not have to

undergo the required system safety analysis (which may, in the end, still not have picked up on the problem of wear and the risks it implied). The distinction, however, shows that airworthiness is not a rational product of engineering calculation. Certification can have much more to do with localized engineering judgments, with argument and persuasion, with discourse and renaming, with the translation of numbers into opinion, and opinion into numbers — all of it based on uncertain knowledge.

Complexity and the limits of measuring, regulating, controlling

As a result, airworthiness is an artificially binary black-or-white verdict (a jet is either airworthy or it is not) that gets imposed on a very grey, vague, uncertain world — a world where the effects of releasing a new technology into actual operational life are surprisingly unpredictable and incalculable. The term "unruly technology" was introduced by Brian Wynne in 1988 to capture the gap between our image of tidiness and control over technology through design, certification, regulation, procedures, maintenance, on the one hand, and the messy, not-so-governable interior of that technology as it behaves when released into a field of practice. Technology and safety is about how things behave in context, not on the drawing board. Universal proclamations or assurances about reliability figures are mute when ideas or designs are put into working solutions in this or that situation. A crucial skill involves finding a practical balance between universality of safety assumptions and their contextualization (Wynne, 1988).

This is a balance (and a possible gap) that not only operates between those on the outside (the public, or consumers of the technology) and insiders (engineers, managers, regulators), but applies even to insiders themselves: to practitioners very close to the technology. If the operational system is not itself following the rules by which it was predicted or supposed to operate, insiders can also reconcile such data with their beliefs — not by changing beliefs, but by looking differently at the data. This works particularly when (1) it is a relatively common problem, among a mass of other relatively common problems, (2) there are routine operational ways of compensating for it (providing more redundancy), and (3) alternative approaches would severely disrupt the economic or operational viability of the system (Wynne, 1988). The stretching of maintenance intervals, even against the background of a technology that behaves differently in practice than what was predicted, is a relatively common issue. Through these processes, insiders can keep their beliefs intact, they can retain their image of tidiness and controllability. But the technology may remain unruly, no matter what.

With as much lubrication of the jackscrew assembly as it originally recommended, Douglas thought it had no reason to worry about thread wear. So before 1967, the

manufacturer provided or recommended no check of the wear of the jackscrew assembly. The trim system was supposed to accumulate 30,000 flight hours before it would need replacement. But operational experience revealed a different picture. After only a year of DC-9 flying, Douglas received reports of thread wear significantly in excess of what had been predicted. The technology, in other words, refused to play by the manufacturer's rules. In response, the manufacturer recommended that operators perform a so-called end-play check on the jackscrew assembly at every maintenance C check, or every 3,600 flight hours. The end-play check uses a restraining fixture that puts pressure on the jackscrew assembly, simulating the aerodynamic load during normal flight. The amount of play between nut and screw, gauged in thousandths of an inch, can then be read off an instrument. The play is a direct measure of the amount of thread wear.

From 1985 onward, coinciding with the deregulation of the airline industry, end-play checks at Alaska Airlines became subject to the same kind of drift as the lubrication intervals. In 1985, end-play checks were scheduled every other C check, as the required C checks consistently came in around 2,500 hours, which was rather ahead of the recommended 3,600 flight hours. This would unnecessarily ground the aircraft for maintenance that was not officially due yet. By scheduling an end-play test every other C check, though, the interval was extended to 5,000 hours. By 1988, C-check intervals themselves were extended to 13 months, with no accompanying flight-hour limit. End-play checks were now performed every 26 months, or about every 6,400 flight hours. In 1996, C-check intervals were extended once again, this time to 15 months. This stretched the flight hours between end-play tests to about 9,550.

The last end-play check of the accident airplane was conducted at the airline maintenance facility in Oakland, California in 1997. At that time, play between nut and screw was found to be exactly at the allowable limit of .040 inches. This introduced considerable uncertainty. With play at the allowable limit, what to do? The rules were not clear. The so-called AOL 9-48A said that "jackscrew assemblies could remain in service as long as the end-play measurement remained within the tolerances (between 0.003 and 0.040 inch)" (NTSB, 2002). It was still 0.040 inches, so the aircraft could technically remain in service. Or could it? How quickly would the thread wear from there on? Six days, several shift changes and another, more favorable end-play check later, the airplane was released. No parts were replaced: They were not even in stock in Oakland. The airplane "departed 0300 local time. So far so good," the graveyard shift turnover plan noted (NTSB, 2002).

Three years later, the trim system snapped and the aircraft disappeared into the ocean not far away. Each extension of the interval made local sense, and was only an increment away from the previously established norm. No rules were violated, no laws broken. Even the regulator concurred with the changes in end-play check intervals. These were normal people

doing normal work around seemingly normal, simple, stable technology. Even the manufacturer had expressed no interest in seeing these numbers or the slow, steady degeneration they may have revealed. If there was drift, in other words, no institutional or organizational memory would know it. But isn't that exactly what we build protective structures for? Regulators, maintenance programs, accountable managers, nominated post-holders?

Failure brews where it is prevented

The paradox of complex, safety-critical systems, is that the potential for accidents brews non-randomly, opportunistically, in precisely the structures and processes of governance and organization that are supposed to prevent the accident (Pidgeon & O'Leary, 2000). As the whole story of MSG's and regulators and maintenance organizations and airlines shows, there is a large web of innumerable relationships in which the operation of risky technology is suspended — a web, moreover, that has no clear boundaries, no obvious end or beginning. This structure, which is designed (and evolved) to keep a technology safe, can make the functioning and malfunctioning of that technology more opaque (Clarke & Perrow, 1996). The meaning of signals about the technology (e.g. that it is not behaving according to original manufacturer specifications, or that people are not abiding by the latest procedure) get constructed, negotiated, and transacted through the web of relationships that is strung throughout this structure. The weak signals that are left over trigger only weak organizational responses, if any at all (Weick & Sutcliffe, 2001). The protective structure itself contributes to the construction and treatment of weak signals in ways that are inadvertent, unforeseen, and hard to detect.

The organized social complexity surrounding the technological operation, all the maintenance committees, working groups, regulatory interventions, approvals, and manufacturer inputs, that all intended to protect the system from breakdown, could actually help set its course to, and over, the edge of the envelope. Take the lengthy, multiple processes by which maintenance guidance was produced for the DC-9 and later the MD-80 series aircraft. Alaska 261 illustrates the large gap between the production of a system and its operation. Inklings of that gap appeared in observations of jackscrew wear that was higher than what the manufacturer expected. Not long after the certification of the DC-9, people began work to try to bridge the gap. Assembling people from across the industry, a Maintenance Guidance Steering Group (MSG) was set up to develop guidance documentation for maintaining large transport aircraft, particularly the Boeing 747. Using this experience, another MSG developed a new guidance document in 1970, called MSG-2, which was

intended to present a means for developing a maintenance program acceptable to the regulator, the operator, and the manufacturer.

The many discussions, negotiations, and inter-organizational collaborations underlying the development of an "acceptable maintenance program" showed that how to maintain a once-certified piece of complex technology was not at all a solved problem. In fact, it was very much an emerging understanding, open to constant revision and negotiation. Technology that appeared simple and certain on the drawing board, proved more unruly. It was not before it hit the field of practice that deficiencies became apparent — if one knew where to look.

In 1980, through combined efforts of the regulator, trade and industry groups and manufacturers of both aircraft and engines in the US as well as Europe, a third guidance document was produced, called MSG-3. This document had to deconfound earlier confusions, for example, between "hard-time" maintenance, "on-condition" maintenance, "condition-monitoring" maintenance, and "overhaul" maintenance. Revisions to MSG-3 were issued in 1988 and 1993. The MSG guidance documents and their revisions were accepted by the regulators, and used by so-called Maintenance Review Boards (MRB) that convene to develop guidance for specific aircraft models.

A Maintenance Review Board, or MRB, does not write guidance itself, however; this is done by industry steering committees, often headed by a regulator. These committees in turn direct various working groups. Through all of this, so-called on-aircraft maintenance planning (OAMP) documents get produced, as well as generic task cards that outline specific maintenance jobs. Both the lubrication interval and the end-play check for MD-80 trim jackscrews were the constantly changing products of these evolving webs of relationships between manufacturers, regulators, trade groups, and operators, who were operating off of continuously renewed operational experience, and a perpetually incomplete knowledge base about the still uncertain technology (remember, end-play test results were not recorded or tracked).

The introduction of a new piece of technology is followed by negotiation, by discovery, by the creation of new relationships and rationalities. "Technical systems turn into models for themselves," said Weingart, "the observation of their functioning, and especially their malfunctioning, on a real scale is required as a basis for further technical development" (Weingart, 1991, p.8-9). Rules and standards do not exist as unequivocal, aboriginal markers against a tide of incoming operational data (and if they do, they are quickly proven useless or out of date). Rather, rules and standards are the constantly updated products of the processes of conciliation, of give and take, of the detection and rationalization of new data. As Wynne said:

> Beneath a public image of rule-following behavior and the associated belief that accidents are due to deviation from those clear rules, experts are operating with far greater levels of ambiguity, needing to make expert judgments in less than clearly structured situations. The key point is that their judgments are not normally of a kind—how do we design, operate and maintain the system according to 'the' rules? Practices do not follow rules, rather, rules follow evolving practices (Wynne, 1988, p. 153).

Nor is there a one-way and unproblematic relationship between the original rules or requirements and subsequent operational data. Even if the data, in one reading, may prove the original requirements or rules wrong, this doesn't mean that complex systems, under the various normal pressures of operating economically, reject the requirements and rules and come up with new ones. Instead, the meaning of the data can get renegotiated. People may want to wait for more data. The data can be denied. Or the data can be said to belong to a category that has nothing to do with safety but rather with normal operational variance (CAIB, 2003; Vaughan, 1996).

Conclusion

The very nature of complexity makes any project to prevent future accidents quite daunting. Whereas Systems Thinking 1.0 has a clear agenda for action in investigation and prevention (find the holes, fix the broken parts), Systems Thinking 2.0 has much less clarity to offer. As Pidgeon and O'Leary point out:

> […] we can ask whether our analyses and theories of past accidents and disasters tell us anything useful at all for designing institutions with better future performance, or whether we are merely left with the observation that complex organizations, faced with turbulent environments, will repeatedly fail us in unpredictable ways (and that the only practical advice to risk managers is to stay fully alert to this possibility)? (Pidgeon & O'Leary, 2000)

Complexity happens even if we don't want it to happen. Yet speaking the language of complexity can help us find leverage points. In a complex system, an action controls almost nothing. But it influences almost everything (Page, 2008). Most managers, just like investigators, have operated with a machine model of the world, thinking they can control things, or (in case of investigators) that some people at some point in time could or should have controlled everything. This model is predicated on linear thinking, on symmetry between causes and effects, on predictability. And it is predicated on control.

The commitment that is called for in Systems Thinking 2.0 is to see safety-critical organizations as complex adaptive systems. In those, order is not easy to impose through control from above. In complex systems order arises because of (or emerges from) the interaction of lower-order components and their interaction with their environment. It is not

easy to predict in detail how emergent order will look, but it is easy to predict that some kind of order will emerge. Managers should always expect surprises, no matter how simple their goal may seem. Decisions may get dampened along the way, but they may also get amplified, reproduced, copied. Through positive feedback loops, decisions may produce more similar decisions, which produces even more similar decisions later and elsewhere. That is the sort of path-dependence that can help produce accidents. In complex systems the effects of local decisions seldom stay local. In a tightly interconnected and interdependent world, decisions have influences way beyond the knowledge or computational capacity of the one who makes them.

Insights from resilience engineering and complexity science all point to the importance of diversity. Resilience in a complex system is the ability to recognize, adapt to and absorb problem disturbances without noticeable or consequential decrements in performance (Hollnagel, 2006). Diversity is a critical ingredient for resilience, because it gives a system the requisite variety that allows it to respond to disturbances. With diversity, a system has a larger number of perspectives to view a problem with and a larger repertoire of possible responses. Diversity means that routine scripts and learned responses do not get over-rehearsed and over-applied, but that an organization has different ways of dealing with situations and a richer store of perspectives and narratives to interpret them with.

Systems that don't exhibit diversity will be driven to pure exploitation of what they already know. Little else will be explored and nothing new will be learned; existing knowledge will be used to drive through decisions. This has been called a take-over by dominant logic, or group think (Janis, 1982). One of the positive feedback loops that starts working with these phenomena is selection. People who adhere to the dominant logic, or who are really good at expressing the priorities and preferences of the organization in how it balances production and risk, will excel and get promoted. This creates, reproduces and legitimates an upper management that believes in the dominant logic, which offers even more incentives for subordinates to adhere to it as well.

There is, of course, need for a balance in the promotion of diversity. Too much diversity can mean that the system will keep on exploring new options and courses of action, and never actually settle on one that exploits what it has already discovered and learned. In time-critical situations, decisions may have to be taken without extensive exploration. Rather, the knowledge that is available right there and then must be exploited even though better alternatives may lie just around the corner in the future.

References

CAIB. *Report Volume 1, August 2003*. Washington, DC: Columbia Accident Investigation Board, 2003.

CILLIERS, P. *Why we cannot know complex things completely*. *Emergence, 4*(1/2), 2002, p. 77-84.

CILLIERS, P. Complexity, deconstruction and relativism. *Theory, Culture & Society, 22*(5), 2005, p. 255-267.

CLARKE, L.; PERROW, C. Prosaic organizational failure. *American Behavioral Scientist, 39*(8), 1996, p. 1040-1057.

HOLLAND, J. H. *Emergence*. Reading, MA: Addison-WesleY, 1998.

HOLLNAGEL, E. (2006). Resilience: The challenge of the unstable. In E. HOLLNAGEL; D. D. WOODS; LEVENSON, N. (Eds.), *Resilience Engineering: Concepts and Precepts*. Aldershot: Ashgate Publishing Co., 2006.

HOLLNAGEL, E.. *The ETTO Principle: Efficiency-Thoroughness Trade-Off. Why things that go right sometimes go wrong*. Aldershot, UK: Ashgate Publishing Co., 2009.

Janis, I. L. *Groupthink, Second Edition*. Chicago, IL: Houghton Mifflin, 1992.

LEVESON, N. G. Applying systems thinking to analyze and learn from accidents. *Safety Science, 49*(1), 2011, p. 55-64.

NTSB. *Loss of control and impact with Pacific Ocean, Alaska Airlines Flight 261 McDonnell Douglas MD-83, N963AS, about 2.7 miles north of Anacapa Island, California, January 31, 2000* (No. AAR-02/01). Washington, DC: National Transportation Safety Board.

PAGE, S. E. (2008). Uncertainty, difficulty and complexity. *Journal of Theoretical Politics, 20*(2), 2002, p. 115-149.

PIDGEON, N.; O'LEARY, M. Man-made disasters: why technology and organizations (sometimes) fail. *Safety Science, 34*(1-3), 2000, p. 15-30.

POOLE, R. W.; BUTLER, V. Airline deregulation: The unfinished revolution. *Regulation, 22*(1), 8, 1999.

RASMUSSEN, J. Risk management in a dynamic society: A modelling problem. *Safety Science, 27*(2-3), 1997, p. 183-213.

ROESLER, A.; FEIL, M.; PUSKEILLER, A.; WOODS, D. D.; TINAPPLE, D. *Design is telling stories about the future*. Columbus, OH: Cognitive Systems Engineering Laboratory, The Ohio State University 2001.

STARBUCK, W. H.; MILLIKEN, F. J. Challenger: Fine-Tuning the Odds Until Something Breaks. *The Journal of Management Studies, 25*(4), 1988, p. 319-341.

TURNER, B. A. *Man-made disasters*. London: Wykeham Publications, 1978.

VAUGHAN, D. *The Challenger launch decision: Risky technology, culture, and deviance at NASA*. Chicago: University of Chicago Press, 1996.