Anticipating Adverse Black Swan Events

Michael Barwise, The Risk Studio

"It is the view of the Columbia Accident Investigation Board that the Columbia accident is not a random event, but rather a product of the Space Shuttle Program's history and current management processes."[1]

The Black Swan

Major disasters are now frequently described as Black Swan events, but does this really mean something, or is it just another journalistic cliché? The currently accepted definition of Black Swan Events comes from Nassim Taleb's influential 2007 book "The Black Swan"[2] in which he attempted to formalise and extend a loose concept that has been around since at least the time of Juvenal (1st century CE) - originally, that of a phenomenon so unlikely as to be effectively impossible. Taleb took the idea a stage further, defining a Black Swan event as one that *"...lies outside the realm of regular expectations, because nothing in the past can convincingly point to its possibility. Second, it carries an extreme impact. Third, in spite of its outlier status, human nature makes us concoct explanations for its occurrence after the fact, making it explainable and predictable."*[3]

Even supposing we accept the general premise, this doesn't help us discover what we really need to know - how to anticipate a Black Swan event. It only describes what one looks like from a certain external viewpoint. Risk management guru Dr. Sally Leivesley has asserted that *"Extreme risk cannot be managed by probability theory, statistics or historical loss data"*[4] which helps a bit by alerting us some potential false starts, but we're still substantially in the dark - we now know a Black Swan event is extreme and has random properties, but for the rest we've only been told what it's not.

Nevertheless Taleb's definition has been subsumed into the risk management vocabulary and is bandied about with enthusiasm. My concern is that, while seductive, it doesn't help us very much in actually managing risk.

Avoiding Pessimism

The 9/11 terrorist attacks, conflicts in oil-producing nations and the Fukushima nuclear disaster are among the many that have been described as Black Swan Events, and the concept was even voiced in anticipation of the final flight of the NASA space shuttle in July 2011. But not every extreme occurrence is necessarily a Black Swan. So how does one qualify? According to Taleb's definition it has to have an extreme outcome and be difficult to predict, but I assert that it also has to be causally complex - that's the prime source of its unpredictability. Some extreme events are outcomes of very simple causal sets, but many others take us by surprise because the totality of their evolution looks unlike anything we have encountered before. Not necessarily the nature of the outcome, but - essentially - the total pattern of contributory factors leading up to it, appears unique.

The impotence of conventional statistical techniques in the face of this apparent uniqueness doesn't however mean such events are inherently impossible to anticipate - and therefore to take precautions against. And unless we're mere armchair philosophers, that is of course our real task. To fulfil it, we have to start from an appropriate frame of reference, so we must first distinguish between prediction and anticipation. Prediction implies some notion of a nominal event frequency or a time scale within which the event can be expected to occur. That's often uncertain even where there's a substantial body of comparable prior events, and indeed probability theory itself doesn't promise so much for rare events.

The required body of evidence appears at first sight to be absent in the case of Black Swan events, as each is in its totality unique by definition. So Taleb's emphasis on the unpredictability of a Black Swan event really does little more than beg the question, leaving us as helpless as he implies. But I believe his assertion that *"nothing in the past can convincingly point to its possibility"* is untenable. Although a Black Swan event's causal complexity makes them difficult to unravel from the whole, many of its contributory components will assuredly have recognisable precursors - often too many for comfort. The real question is whether we're able to identify them and recognise their significance as contributors to the whole in advance. But that depends as much on our powers of observation as on the nature of the factors themselves. If we are able to identify them sufficiently clearly, we may well find there are recognisable commonalities between the causal clusters of Black Swan events that make them only - to use the salesman's term - "almost unique". Which means Black Swans may actually be both much more common and less black than we thought. So although we cannot with any assurance predict when such an event will occur, we might confidently declare it to be inevitable, based on an appropriate assessment of a cluster of individually identifiable potential causal factors. If we can indeed do this, we will have the opportunity to reduce the likelihood of the event occurring by exercising precautionary measures against its causal factors.

Thinking about Risk

Although Taleb's proposition that the thought process itself is our biggest problem is valid, his specific conclusion - *"We do not spontaneously learn that we don't learn that we don't learn..."*[5] is not. Neither failing to learn that we don't learn nor indeed failing to learn is our biggest problem. We do learn a great deal, although we don't always learn what is objectively most important. But where his argument fundamentally collapses is the tacit assumption of immutability - *"The problem lies in the structure of our minds"*. That is refuted by a large and growing body of psychological research. It seems instead that in our superficially fast-moving urban-commercial cultures there's commonly a lag between the adaptation of our thought processes and the changing demands placed upon them - loosely expressed, we spend much of our time outrunning materially the evolution of our mental processes. In the words of neurophysiologist Robert Ornstein *"the survival problems we now face are ... complex, if only because they exist at an emergent level of thought and action"*[6] and we have trouble keeping up with the resultant changes that both surround and are required of us.

But this is more a result of how we are trained to apply our mental faculties than of any intrinsic limits. The Western intellectual tradition almost exclusively emphasises sequential rational and analytic cognition that engages only a small part of our mental capacities, at the expense of holistic perception and interpretation. Nevertheless we have vastly more intrinsic mental plasticity and perceptual capacity than we use in everyday life - or even in conventional risk assessment. For example, in 1987 the Los Angeles Times reported a public experiment[7] in which an audiophile successfully identified the music on unlabelled gramophone records by visually examining their grooves; in his previous incarnation as a record producer, neuroscientist Daniel Levitin learned to identify the brand of tape on which recordings were made by their sound alone,[8] and it took Rockefeller University psychologist William Hirst and his colleagues only a few weeks to train some perfectly ordinary people to read and write independent texts simultaneously with full comprehension of both. Hirst concluded that *"A skilled individual has learned to detect new stimulus constellations and execute new patterns of action, not just to do old things ... unconsciously."*[9]

So I suggest the assumption that a Black Swan event is impossible to anticipate *per se* - either because it's utterly unprecedented or due to our inherent mental incapacity to foresee it - is invalid, and if we subscribe to it we create a self-fulfilling prophecy leading to failure.

We do not have to fail. But the apparent stability and security of our day-to-day lives makes the mental plasticity and perceptual capacity we must employ in order to succeed largely dormant in most of us most of the time. They must be actively wakened and specially trained to detect Hirst's "new stimulus constellations." We have to learn to exercise more acute perception and greater discrimination, and to handle more complex webs of interaction than practically any of the tasks we currently undertake require. And despite Taleb's assertion to the contrary, a good starting point for learning to do this is the objective detailed analysis of the causal clusters of past extreme events. So let's look at such an event and try to learn from it.

Germany, July 1st 2002

Just after 21:35 on July 1st 2002, a Tupolev TU154M out of Moscow collided in mid-air with a Boeing B757-200 transport from Bahrain over Ueberlingen, Germany in Zurich-controlled air space.[10] Both planes crashed with no survivors. Both crews were highly experienced and well rested prior to their respective flights, and all flight systems were apparently fully operational. So what went wrong on the night of July 1st?

In the Air

First, let's follow the time line.[11] Just before 21:22 the Boeing transport with only two crew on board entered Zurich air space. At 21:21:56 it was instructed by Air Traffic Control Zurich (ATCZ) to turn on its transponder - a device that allows it to be identified as a specific plane by radar and by TCAS (the onboard automated collision avoidance warning system) - and to climb from flight level (FL) 260 to 360 (from 26,000 feet to 36,000 feet). Its climb was completed by 21:29:50.

At 21:30:11 the Tupolev, already flying at FL 360 - with a pilot flying, an off-duty co-pilot, an instructor operating communications, and two other crew on the flight deck - called in to ATCZ and was told to turn on its transponder. About three minutes later, a discussion started on the flight deck of the Tupolev about some other aircraft at the same altitude that had appeared on their instruments.

At 21:34:24 the co-pilot of the Boeing went to the toilet, leaving only the pilot on the flight deck.

At 21:34:36 the Tupolev pilot flying said *"Here it is in sight."* Six seconds later, the TCAS on both planes issued a *"traffic traffic"* alert indicating another aircraft nearby at the same altitude. The Tupolev co-pilot echoed the *"traffic traffic"* warning verbally.

At 21:34:49 ATCZ issued an instruction to the Tupolev to descend smartly to FL 350 due to conflicting traffic. The Tupolev commenced its descent at 21:34:54, but didn't acknowledge to ATCZ.

At 21:34:59 the TCAS on both planes issued a manoeuvring alert requiring immediate action due to close proximity between them - that on the Boeing instructing it to descend and that on the Tupolev to climb. The Tupolev co-pilot said *"It says 'climb'."* The instructor (not flying) replied *"He is guiding us down."* The co-pilot queried *"descend?"*

The Boeing crew obeyed their TCAS and descended, but the Tupolev pilot flying also continued to descend, ignoring the TCAS and following the previous instruction from ATCZ.

The planes were now less than seven nautical miles apart on intersecting courses at a combined air speed in the order of 700 knots - about half a minute from interception in the absence of avoidance manoeuvres.

At 21:35:03, fourteen seconds after its first *"descend"* instruction, ATCZ again ordered the Tupolev to descend smartly and to acknowledge. ATCZ then immediately followed with *"traffic at your 2 o'clock position at 3-6-0."* The Boeing was in fact at 10 o'clock and already at FL 356. The Tupolev co-pilot spotted it immediately and the navigator said *"it's going to pass beneath us."* The pilot of the Tupolev increased his rate of descent to more than 2,000 feet per minute.

At 21:35:10 the Boeing TCAS issued an *"increase descent"* alert. The pilot complied, around which time the Boeing co-pilot returned to his seat. Nine seconds later the Boeing reported its TCAS-instructed descent to ATCZ.

At 21:35:24 the Tupolev TCAS issued an *"increase climb"* alert. The co-pilot commented *"It says 'climb'."* The plane's rate of descent was then around 1,800 feet per minute.

At 21:35:26 the Boeing co-pilot said *"descend."* The plane was by this time already descending at about 2,600 feet per minute.

At 21:35:30 the Boeing co-pilot urgently commanded *"descend [expletive] hard"*, and the pilot pushed the control column fully forward.

The collision occurred two seconds later at 34,890 feet, the Boeing passing immediately beneath the Tupolev at right angles slicing its fuselage in two with its tail fin, which was ripped off in the process. The Tupolev was still descending at the time of impact, its pilot having only initiated a climb five seconds earlier. One second (approximately 350 metres) before impact, the Tupolev's control column had been pulled right back and the thrust levers pushed fully forward.

Now that, confused as it seems, is technically what happened in the air. But it's only half of even that half of the story. On this flight, the Tupolev's flight crew included an instructor (the chief pilot of the airline) and an off-duty co-pilot with no active duties. All of them except for the instructor formed a fixed crew that had flown together for some time. The instructor's official mission was to supervise the (otherwise highly experienced) active pilot's approach into Barcelona, so he could be cleared for solo approach to that airport - considered a difficult one. But despite not being the "pilot flying" and being only responsible for communications during the flight prior to approach, the instructor seems to have been considered "pilot in command". The off-duty co-pilot had no apparent official standing on this flight. This situation seems to have been accompanied by some tension on the flight deck. The presence of the instructor appears to have disrupted a normally quite rigid hierarchy: the instructor - acting as "pilot in command" - becoming increasingly autocratic as the emergency evolved.

The two-man crew of the Boeing were an established team that seem to have co-operated readily, and the pilot appeared decisive without asserting authority.

Both crews attempted avoidance manoeuvres immediately before the collision, but there was insufficient time for the planes to respond. But on both flight decks it appears that the pilot flying was less in touch with the situation than other members of the crew, particularly in the final half minute.

On the Ground

Now let's take a look at events on the ground. On the night of July 1st, things were not quite normal at Air Traffic Control Zurich. At 21:00, thirty-five minutes before the accident occurred, a six-hour maintenance was scheduled to start on the flight plan processing systems. Once the maintenance had commenced, half a dozen technicians were working in the control room in close proximity to the controllers and their assistants on duty. Several of the radar systems were affected, as was the primary ground-to-ground telephone system, which had to be switched off entirely. The controllers on that night shift had not been warned of the maintenance when they came on duty at 17:50, although documents describing the work were available to be

read. The controllers did not refer to them, but it would not have helped very much to have done so, as they did not describe the effects of the work on the availability of traffic control systems.

The switch-off of the primary phone system eliminated all dedicated direct telephony, requiring fall-back to a backup system using the public telephone service to route all calls. The backup had a pre-programmed list of outgoing numbers and it was not possible to call other numbers. Incoming calls directed to the primary system could not be transferred to the backup system. The shutdown of the primary system was omitted from the maintenance notices provided to ATCZ, and although it was technically available again by 21:34:37 - almost a minute before the crash and about the time the first TCAS alert occurred - the controller was not given clearance to use it. Between 21:34:44 and the moment of impact three calls from the Karlsruhe upper control area and one from Friedrichshafen ATC came in on the now-restored primary phone system but went unanswered. For some unspecified period encompassing around 21:30 and the moment of impact, the backup phone system exhibited a fault that prevented the ATCZ controller contacting Friedrichshafen.

There was apparently an unofficial convention at ATCZ that once traffic reduced on the night shift, one of the two controllers could retire to a lounge, to be recalled when the traffic volume increased again. But the lounge was distant from the control room, which could only contact the resting controller by telephone. The resting controller left for the lounge at 21:15 that night, and the resting controller's assistant followed ten minutes later, just as events were about to unfold. So by the time the incident began to evolve, a single controller was in charge of all air traffic at ATCZ, having already been on shift for about three hours.

This sole controller simultaneously had to monitor two control stations nearly two metres apart - one covering the whole Zurich air space and the other a narrower area - and was acting as chief controller on duty as well. He was apparently experienced at single-handed operation on the night shift under normal circumstances. But that night radar was intermittent due to the maintenance operations and in the final moments the controller was handling planes on both stations simultaneously - an incipient, if not fully recognised, collision on one and a late arrival Airbus A320 on approach on the other. His assistant, although present, was not rated for air traffic control duties. Communications between the two closing aircraft and ATCZ were disrupted by simultaneous transmissions from the incoming Airbus A320 on a different frequency.

Discussions I have had with experienced pilots indicate that once TCAS has issued a warning, it should take priority over ATC instructions until the situation normalises. Indeed as soon as TCAS engages between two planes, ATC is supposed to back off. But ATC can't tell that a plane's TCAS has gone into action unless the flight crew report it by radio. TCAS typically provides a maximum 50 second window for avoidance from its first manoeuvring advisory, and voice radio communications typically involve a round trip delay of about ten seconds.

Outcomes and Implications

The above is only a summary of the main findings - the investigators' analysis[10] runs to 36 pages. But the bottom line is that two aircraft were destroyed, 71 lives were lost, and the accident investigators of the BFU (German Federal Bureau of Aircraft Accidents Investigation) finally made nineteen separate safety recommendations. Nevertheless there was more to come. The flight controller on active duty that night was murdered two years later by the father of one of the passengers on the Tupolev. That's the unofficial final tail feather of this authentic Black Swan.

According to the record, the incident took just under 14 minutes to evolve from the time the Boeing first called into Air Traffic Control Zurich. But some of the seeds of it were quite possibly sown much earlier - maybe when the Tupolev TU154M flight manual was written - describing as it did the urgent *"climb climb"* alert from the TCAS as a *"recommendation to the crew"*. TCAS is a fully automated system that communicates between planes in close proximity and issues co-ordinated instructions to both to separate them with a few tens of seconds to spare. But in the English language manuals the alerts it generates are also only referred to as *"advisories"*. Or maybe some seeds were even sown long before that, when the TCAS itself was designed without any automatic means of informing air traffic control of its alerts. Or possibly when a backup telephone system was implemented at ATCZ that couldn't interoperate with the primary system. Or indeed when any of the other contributory prior events was set in motion. But it's pretty certain that in the absence of any single one of these factors, the collision might well not have occurred - a near miss maybe, but no more.

From a Different Perspective

So why *did* it happen? It's impossible to identify a single dominant contributor to the total event. Everything - from the way TCAS works and the wording of the Tupolev flight manual to the several technical effects of the ATCZ maintenance; from the design of the control room systems and shortage of controllers to the social

dynamics on the flight decks - had its necessary part in the evolution of the accident. Even the Boeing co-pilot's trip to the toilet was identified as a potential contributor to the outcome, as he was absent from the flight deck when the first TCAS alert occurred and was therefore unable for almost a minute to perform his designated duty of visual search for the other aircraft.

There were so many contributory factors here that it seems at first sight as if we couldn't possibly have covered all the bases. So trying to hypothesise that precise outcome from a list of those factors in advance would be a mug's game. But if we examine things from another direction, they look a bit different.

For rare extreme events the significance of individual component factors is frequently not obvious prior to the event. The Black Swan emerges from hiding only when multiple independent factors coincide in a particular sequence. But, as is clear from the Ueberlingen incident, efforts by those involved to contain a developing incident can actually lead to exacerbation due to suboptimal decisions made under growing pressure. So prevention is better than relying on attempts to cure.

Prevention is best approached by mitigation of as many as possible of the potential contributory factors well in advance. And each of these contributory factors, while on its own unlikely to result in an extreme event, can usually be identified as a real albeit lesser hazard in its own right.

But there's a danger when conducting risk assessments that such individual factors get ranked as low risk on their apparent merits in isolation, without sufficient appreciation of their potential contribution to some Black Swan event or other. I say "some ... or other" advisedly, because it's much more realistic in advance to consider extreme events in generic terms than in specific. Generic adverse outcomes are easier to envisage and develop successful countermeasures against than the ultra-detailed hypothetical event that almost certainly won't turn out quite like you thought it would.

The Black Swan Redefined

It's clearly not the case that "*nothing in the past"* could have convincingly pointed to the possibility of the Ueberlingen air disaster. Its causal cluster was replete with "accidents waiting to happen" at many levels and mid-air collisions are a recognised type of aviation disaster. So it fails Taleb's first test. But I assert that it was nevertheless a classic Black Swan event.

As adverse Black Swan events occur due to the chance coincidence of multiple independent less extreme events that combine to create a critical situation, the probability of the event is not computable in advance with any useful degree of confidence. But I think it's safe to say that absent a small proportion of those factors the specific Black Swan event would almost certainly not occur. And most of its component factors can be identified individually as hazards to some degree in their own right - provided we're looking at them from an appropriate viewpoint. But if, as was established by the investigations into both the Challenger and Columbia space shuttle disasters, we habitually downgrade the subjective risk ranking of a perceived hazard each time we get away with disregarding it, the ultimate outcome will assuredly be a Black Swan.

Many of these hazardous factors are likely to be systemic failures - persistent process and systems flaws that individually slip past incomplete analysis or over-confident risk assessment. And although they may originate anywhere in the design, deployment or operational phase, most of them are probably in place long before they combine to yield an extreme outcome. But because it's a "series coupled system" in engineering terms - relying on all the contributory factors to occur in due sequence - a Black Swan event is probably quite fragile, at least in the early stages of its evolution. Any change in the composition of the causal cluster will modify the ultimate outcome. So mitigating even a few of the potential contributory factors can yield a disproportionately great improvement in safety.

From the opposite perspective, the extent and complexity of the technological base we rely on in our highly mechanised societies makes it quite possible that Black Swan events are on the verge of occurring very frequently - failing to materialise only because some individual link or other in the chain of causality is absent or out of the required sequence. It's therefore a high priority to address the components of such chains of causality at source.

So instead of relying primarily on fire fighting - arbitrary "feel good" countermeasures to poorly defined but complicated hypothetical threat scenarios ranging from unjustified assumptions concerning potential targets of attack[12] to the much quoted but fundamentally flawed "ticking bomb" thought experiment [13] or chasing our tails to develop precisely mating responses to specific perceived threat complexes as they emerge[14] - our currently preferred responses to both "Homeland Security" and the protection of the critical electronic infrastructure - we should aim for good process design and systems engineering that intrinsically support resilience and facilitate compliance with commonsense precautions *ab initio*. We don't have to cover *all* the bases in advance, just the self-evident ones - even if not necessarily "extreme" in their own right. I maintain

that practically all potential contributory factors to Black Swan events should be identifiable as non-conformities - with impartial common sense at least - in their own right in advance, provided we pay sufficient attention in the right way. But to recognise them we must both avoid Taleb's pessimism and pay the right kind of attention. Instead of throwing in the towel and dismissing the anticipation of extreme events as "impossible" we must both adopt a much more holistic viewpoint than we are used to and avoid preconceptions based on wishful thinking or apparent expediency. As "Crow's Law"[15] states *"Do not think what you want to think before you know what you ought to know."*

So to sum up, I'd like to offer an alternative - and I hope both a more precise and a more useful - definition: *"A Black Swan event is an extreme outcome resulting from the chance coincidence of a large number of independent factors that, when considered independently in advance, are not expected to coincide to yield that outcome, given the way we think about them."*

The key is of course *"given the way we think about them"*. The mind set we typically bring to risk assessment at present doesn't seem very effective when applied to extreme events, and Taleb's definition of the Black Swan seems to me merely a justification for our failure. Clearly, we're not thinking about extreme risk in the right way - if we were there would be fewer incidents. But although entrenched, our thought processes are not set in stone. We can, and must, learn to employ a more holistic and impartial perspective. So it's here that our attention should be focused - not on the complex detail of some hypothetical Black Swan, nor even in the first instance on the specifics of potential component events, but before all on the thought processes we use to identify, evaluate and mitigate them. In the words of the late Owen Barfield *"There may be times when what is most needed is, not so much a new discovery or a new idea as a different 'slant': I mean a comparatively slight adjustment in our way of looking at the things and ideas on which attention is already fixed."*[16]

I suggest that such a time is now.

References:

[1] Report Volume I, Columbia Accident Investigation Board, August 2003

[2] Taleb, Nassim Nicholas. The Black Swan: The Impact of the Highly Improbable, Allen Lane 2007, ISBN 978-0713999952

[3] Taleb NN. The Black Swan: The Impact of the Highly Improbable, New York Times, 22/04/2007

[4] Leivesley, S. InfoSec as Intellectual Capital in the Business Continuity Paradigm. Information Security Bulletin, 8, 8, October 2003. pp. 311-318

[5] Taleb NN. *supra*

[6] Ornstein, Robert. The Mind Field. London, Octagon Press 1976, ISBN 0-86304-011-X, p. 14

[7] Seckel, Al. The Man Who Could Read the Grooves. Los Angeles Times, October 19, 1987

[8] Levitin, Daniel. This is Your Brain on Music. London, Atlantic Books 2006, ISBN 978 1 84354 716 7, p. 3

[9] Hirst, W *et al*. Dividing Attention Without Alternation or Automaticity. Journal of Experimental Psychology: General, 1980, Vol. 109, No. 1, 98-117

[10] BFU Investigation Report AX001-1-2/02 (May 2004)
http://www.skybrary.aero/bookshelf/content/bookDetails.php?bookId=414

[11] BFU Investigation Report AX001-1-2/02 (May 2004) appendix 2
http://www.skybrary.aero/bookshelf/content/bookDetails.php?bookId=415

[12] http://www.lightbluetouchpaper.org/2006/08/02/security-theater-at-the-grand-coulee-dam/

[13] Adams, Nick et al. Counterterrorism Since 9/11. Oakland, CA, USA, The Breakthrough Institute 2011.
http://thescienceofsecurity.org/blog/CT%20Since%209-11_by_Breakthrough.pdf, p.33

[14] http://reason.com/archives/2009/01/16/safe-but-also-sorry

[15] John Crow, quoted in Jones, RV. Most Secret War. Ware UK, Wordsworth Editions Ltd. 1998, ISBN 1 85326 699 X, pp 76-77.

[16] Barfield, Owen. Saving the Appearances: A Study in Idolatry. London, Faber & Faber 1957

Author:

Michael Barwise is a UK-based policy and risk management consultant with a background in information assurance and systems engineering. He has contributed to international policy and national legislation on Internet and information risk, has lectured on policy development at Masters level, and is widely published online. His special interest is assisting business to improve the quality of risk decision-making.