2009

The Psychology of Aviation Surprise: An 8 Year Update Regarding the Noticing of Black Swans

Christopher D. Wickens

ISAP Keynote

THE PSYCHOLOGY OF AVIATION SURPRISE: An 8 YEAR UPDATE REGARDING THE NOTICING OF BLACK SWANS

Christopher D. Wickens
Alion Science & Technology, Boulder, Colo.
& University of Illinois Human Factors Division

We describe the limitation that people have in noticing very unexpected, surprising "off-nominal", or black swan events, as reflected in the psychology of change blindness; and how this limitation can compromise aviation safety. We then describe a three phase program of research examining pilot response to these black swan events, using (1) a meta-analysis to reveal the miss rate in noticing black swans, (2) a model of visual attention to predict this miss rate, and (3) the same model to make predictions regarding the safety impact of NextGen technology and procedures.

In 2006, an Embraer Legacy business jet and a commercial 737 passenger aircraft collided in mid air over Brazil (Command of Aeronautics, 2006). The 737 was seriously damaged, crashed, and all lives were lost. While, as in any fatal aircraft accident, there were many factors responsible, one of the most critical is that the transponder on the Embraer was not sending its position, such that a TCAS alert in the 737 would have registered the impending collision and an evasive maneuver could have taken place. At an earlier time in the flight history, there is good evidence from air traffic control communications that the Legacy **was** transponding (in response to interrogation) so, at some time prior to the collision, the communications system within the Legacy must have become disabled, and a display within the cockpit changed its state to signal this event, one of great significance and importance, but one that the pilots on board apparently failed to notice. We will emphasize below that this "failure" is one that is quite understandable given the frailties of human attention. We point out here that this provides a prototypical example of the criticality to aviation safety, of noticing unexpected, and often not very salient "off-nominal" events or, to use the term coined by Taleb (2007), "black swan" events.

This issue then is a key element of the "psychology of surprise", which was the focus of my talk at this symposium in 2001 (Wickens, 2001). Since that time, two key elements have led me to revisit this theme, 8 years later. First, at that time, I expressed regret at the lack of much valid data, in realistic environments, that could help aviation psychologists understand pilot (and controller) response to the black swans. A good deal more of such data exist now, and will be summarized below. Second, we are entering a period when revolutionary changes in the airspace are forecast, as reflected by the proposed procedures and equipment that are embodied in the next generation of the airspace or NextGen (JPDO, 2008). Such changes are designed to increase the productivity in the airspace, while preserving levels of safety. Predictive models are being developed to demonstrate the assumed productivity (i.e., capacity) benefits of procedures like merging and spacing, self separation, RNAV, and equivalent visual operations. It is important however that valid models also be developed to predict the safety implications of these productivity enhancements; we argue here that safety concerns **in an already very safe system** must, by definition, be associated with 'black swan" unpredictable events. If such events were predictable, then their consequences would have been mitigated. Any such model will be unlikely to predict **when** such an event will occur, but it should be able to predict the conditions that might make a black swan more possible, as well as key features of the human response to the event, which is the focus of this address.

If we examine the human (e.g., aviation worker) response to very surprising events, we can identify three categories of processes where the response might break down:

1. In noticing (or failing to notice) the triggering event. The Embraer mid-air collision provided such a prototypical example, with the apparent failure to notice the display change signaling the cessation of position broadcast. Another example would be the runway overshoot crash on take-off at Lexington Kentucky (NTSB, 2007), where pilots failed to notice important cues that they had lined up for approach on the wrong runway.
2. In diagnosing. Although airspace workers may notice that things are not right in a timely fashion. (This is, after all, the process that alarm systems support), they may not fully understand the nature of the unexpected problem: a correct situation assessment. Pilots in the CFIT accident near Cali Columbia were aware of a navigational problem, but did not understand, until too late, the course of their trajectory relative to the mountains.
3. In selecting and executing appropriate procedures. A number of aircraft accidents in the previous decade, associated with the flight management system (Dornheim, 1995), were attributed to pilots, "fighting" the autopilot. For example in the Air China crash at Nagoya Japan, the pilot and autopilot were imposing opposite forces on the plane's elevators, until an abrupt pitch up attitude caused a stall.

Of these three stages of pilot information processing, I focus on the first – noticing – for two important reasons. First, it is a well-defined safety bottleneck. Jones and Endsley (1996) have surveyed the literature and found that stage 1 situation awareness breakdowns (which can roughly translate to failure of noticing and/or perception) account for 76% of SA related errors in aviation. Second, such failures directly reflect the psychology of change blindness or inattentional blindness (Simons & Levin, 1997; Rensink, 2002), a striking phenomenon well researched in basic psychological laboratory, that "scales up" remarkably well to applied worlds of driving, flying, and process supervision. (Carpenter, 2001, Martens, 2007; Wickens Thomas & Young, 2000, Sarter, Mumaw & Wickens, 2007, Stelzer & Wickens, 2006).

The phenomenon of change blindness, whereby people are quite insensitive to noticing changes or events in the world around them has three characteristics, and change blindness will be more prevalent to the extent that all three are present:

- The event occurs away from foveal vision, and bear in mind that at any given time, only about 0.02% of the visual world occupies a pilot's foveal vision.
- The event is relatively subtle or non-salient (e.g., not a flashing light, but just the appearance or disappearance of a visual object or displayed element).
- The event is unexpected: Here what is meant by "unexpected" can range between events such as a conflict alert in an ATC facility, which occurs rarely, but the controller assumes that one could happen at any time, and events such as the offset of the transponder signal, for which there may be no expectation whatsoever. We have referred to these as "unexpected" versus "truly surprising" events, and Taleb (2007) has called them gray swans and black swans respectively.

In addition to these three features contributing to change blindness, data suggest that the failure to notice will be amplified as resources are withdrawn from monitoring by dual task conditions, such as those prevalent in the descent phase of flight, particularly in single pilot operations.

How prevalent is change blindness in these circumstances? In a study of cockpit displays of traffic information to support self-separation, Stelzer and Wickens (2006) observed that pilots failed to notice over 80% of changes to flight trajectories of nearby aircraft on the (fairly cluttered) display in front of them, given that they were also engaged in a primary flight task (miles in trail distance keeping), and that the events themselves were not signaled by a warning such as a flash; but rather just a change in a digital data tag (for altitude), or movement direction across the display (for heading). In another study, pilots were inferred to 'miss" around 50% of flight mode annunciator changes in a full mission simulator, at least as this miss rate was inferred from the absence of a visual fixation on the FMA following the change (Sarter et al., 2007). Importantly, in both of these cases, we can refer to the events as gray swans, not black swans, since participants were very much aware that such changes could occur in the context of the experiment.

For extrapolation to aviation safety our research team was more concerned with the actual "black swan" events, for which we suspected that noticing rate might be lower (even though the base rate of such events would also be, by definition, drastically lower). However our challenge was to find **statistically reliable** estimates of such a miss rate, and of the causal effects that could moderate it. The challenge here of course is that by definition, in any experiment, from the perspective of the pilot once such an event occurs once it is no longer totally unexpected (and therefore no longer a black swan). Hence the response to the event can occur only once per pilot per experiment, and this "low N" often thwarts the efforts of researchers to extract statistically reliable data regarding a black swan response.

In order to overcome this challenge to statistical power, in the first of three elements of our research program (Gore et al, 2009), we turned to the technique of meta-analysis (Rosenthal, 1991), in an approach described in detail in Hooey et al. (2009). Here, we identified in the literature every aviation study we could find that used a relatively realistic flight simulation along with licensed pilots, and at some point in the experiment presented a truly surprising, safety-critical black swan event. For example an investigation of synthetic vision systems for landing may present, on the final trial of the experiment, a runway incursion, after several sessions of incursion-free landings (Wickens et al., 2009).

The output of this meta-analysis produced a series of "effects" on off-nominal miss rates that supported our understanding of the safety concern that they engender. We found that overall about 1/3 of the pilots missed these events, and one study (Thomas & Wickens, 2004) was able to attribute such misses in part to pilot scan strategies: those who tended to look less frequently where the event occurred, were less likely to notice it. Importantly, we also found at least four factors that affected this miss rate in a statistically reliable fashion when pooled over studies. Our analyses revealed that pilots had a higher miss rate (MR) for the off nominal event when:

- A black swan outside world event was to be detected while they were flying with a HUD (MR = 0.36) versus without a HUD (MR = 0.27).
- The event was a truly surprising black swan (MR = 0.48) rather than an unexpected gray swan (MR = 0.29).
- The unexpected event occurred down on the instrument panel (MR = 0.39) than out the window ((MR = 0.29)
- An outside-the-cockpit black-swan event occurred while pilots were flying with a head down highway in the sky display (HITS : MR = 0.45) rather than without a HITS (MR = 0.22).
- The off-nominal event was an erroneous clearance delivery and it was delivered by data link alone (MR = 0.69) rather than redundantly with data link and voice (MR = 0.38).

While such relatively low levels of performance might well be considered disconcerting for aviation safety, we also recognize that such misses will occur quite infrequently, since the base rate of these off-nominal black swan events is, by definition, exceedingly low (but not impossible). However one of the ironies of automation (whose failures may often be considered black swan events) is the ironic fact that the rarer the event is, the less expected it becomes, and hence the greater is the likelihood of missing it (Bainbridge, 1983). Furthermore, the results from these high-fidelity flight simulations certainly replicate what is now well known regarding change blindness and inattentional blindness in the real world (Rensink, 2002: Simons & Levin, 1997; Sarter et al., 2007; Stelzer & Wickens, 2007; Wickens & Alexander, 2009; Wickens et al., 2000). That is, people simply do a poor job of noticing changes (events) when these are unexpected, are not salient and occur outside of foveal vision; all conditions that typified the events analyzed in our meta-analysis.

Of course such empirical data as those reported above, while of value in explaining potential concerns with current day (e.g., HUD) and near-future (e.g., HITS) technology, does not inform us of the miss rate of future NEXGEN systems and concepts. For this we must turn to computational modeling (Foyle & Hooey, 2008), which constituted the second element of our 3-element research effort. In this element we developed a model of noticing rare visual events, called N-SEEV. SEEV describes the four components that drive visual scanning around the workplace, typical of the cockpit: Salience, Effort (conservation) Expectancy and Value. Then, in the context of this normal steady-state allocation of visual attention, a to-be-noticed event occurs, in this context the “off-nominal event” (e.g, the onset of a warning signal in the cockpit). Now the noticing (“N”) component of N-SEEV predicts how long the eye will take to land on the location of the event and/or, the probability that it may not be noticed at all (miss rate), or noticed before some deadline.

N-SEEV actually has several parameters (See Wickens et al., 2009); but can effectively drive a simulated eyeball around a simulated cockpit, in a way that does an adequate job of mimicking actual pilot scanning (Sarter et al., 2007) and capturing the variance in noticing time across different display concepts (Nikolic et al., 2004). What we did in the second component of our study was to program N-SEEV to mimic the conditions in which pilots confronted the black or gray swan events, across three of the dichotomous comparisons revealed by out meta analysis: event location, event expectancy (black vs. gray swans) and the presence or absence of a HITS in the cockpit. We correlated the model-predicted miss rate with the observed miss rate from the pilot-in-the-loop simulation (the output of the meta-analysis), and found that our model could predict the actual miss rate of all six conditions in the three contrasts within 14%, and four of the six within 7%.

For those of us interested in the value of computational models in human factors (Foyle & Hooey, 2008) these findings were of great importance because they revealed a good degree of empirical validation of the models in predicting new data. Thus equipped with a model that we believe is valid, the third element of our research was to demonstrate how the model could apply to making predictions of vulnerabilities for certain NextGen procedures and equipment.

For example, these model prediction runs revealed the degree of advantage gained (in noticing in-cockpit warnings) by positioning those warnings close to the primary flight display; but also the high cost, to noticing out-of-the-window black swan events, associated with placing heavy visual demands on the pilot to monitor a CDTI in a self-separation procedure, and particularly with the enhanced cognitive demands when there is a simultaneous engine failure. Here our model predicts that around 80% of these events will be missed.

It is important to reiterate that the likelihood of such occurrences are by definition extremely rare. But, as others have noted (e.g., Taleb, 2007), just because they may be rare, they are not impossible, and so human factors practitioners should be concerned with ways to mitigate these examples of black swan change blindness. In this regard, another advantage of this (and other) computational models emerges: It is a matter of only a few minutes to change parameters of the model (e.g., those associated with an added visual alert, or with re-positioning a display to a HUD location), and a new (and presumably lower) miss rate can be calculated, to signal the safety-advantage of the mitigation.

Of course models are not the panacea for aviation safety. It is important to realize that:

- they are only as good as their validity, and validity itself can only be achieved from empirical pilot-in-the-loop data.

- They can rarely account for all factors (without becoming unwieldy and over-complex). For example N-SEEV only predicts the behavior of a single pilot, and it is unclear how a pair of eyes in the commercial cockpit, with collaborative scan strategies might mitigate some of these effects.
- A model like N-SEEV can rarely predict **when** a miss of a black swan event will occur; or precisely what form the black swan itself will take, but only the circumstances that drive this miss rate up and down and, by extension, how to reduce this rate.

Finally, in conclusion, while we will clearly never be able to eliminate either the occurrence of off-nominal events or the challenge to human attention of noticing them, by understanding what those challenges are, and predicting the circumstances in which they may be amplified, we can go a long way in helping pilots and controllers cope with the “psychology of surprise”.

Acknowledgements

The ideas expressed and data presented in this paper could not have been done without the contributions of my senior collaborators on two NASA research projects: Brian Gore, Becky Hooey, Jason McCarley and Angie Sebok. In addition, Kelly Steelman-Allen, Shaun Hutchins, Julie Bzostek, Corey Koenicke, and Ellen Salud all made vital contributions to model development and data collection. Captain Rick Shay of United Airlines provided vital professional input to the projects. The two NASA projects were NASA’s Airspace Systems – NGATS ATM: Airspace Project (NRA # NX08AE87A) to San Jose State University; subcontracted to Alion Science: MA&D Operations (Dave Foyle, Scientific technical manager) and NASA ROA 2007: Integrated Intelligent Flight Deck Technologies, NNX07AV97A, to the University of Illinois, and also subcontracted to Alion Science: MA&D Operations (Jeff Mulligan, Scientific-technical manager).

References

Carpenter, S. (2001). Sights Unseen. *APA Monitor, 34,* 54-57.

Bainbridge, L (1983). Ironies of Automation. *Automatica, 19,* 775-779.

Command of Aeronautics (2006) final report A-0-22 CENIPA /2008. Brazil. Command of Aeronautics, General Staff of the Aeronautics, Aeronautical Accident Investigation and Prevention Center (CENIPA). (2008). Aeronautical Accident, PR-GTD and N600XL, B-737 8EH and EMB-135 BJ Legacy, 29 September 2006.) Brazilia, Brazil: Cenipa

Dornheim, M. A. (1995, January 30). Dramatic incidents highlight mode problems in cockpit. *Aviation Week and Space Technology*, 57-59.

Foyle, D. C. & B. L. Hooey (Eds.). (2008). *Human performance modeling in aviation*. Boca Raton, FL: Taylor and Francis/CRC Press.

Gore, B.F., Hooey, B.L., , Wickens, C.D., Sebok, Koenecke, C., Salud, E., Hutchins, S & Bzostek, J. (2009a). *Identification of pilot performance parameters for human performance models of off-nominal events in the NextGen environment*. (NASA Final Report and Deliverable. ROA 2006: Airspace Systems – NGATS ATM: Airspace Project, NRA # NNX08AE87A), San Jose, CA: San Jose State University

Hooey, B. L., Wickens, C. D., Salud, E., Sebok, A., Hutchins, S., & Gore, B. F. (2009). Predicting the Unpredictable: Estimating Human Performance Parameters for Off-Nominal Events. *Proceedings* *of the 15th International Symposium of Aviation Psychology*. Dayton, OH: Wright State University.

JPDO (2009). Joint Program & Development Office. Next Generation Air Transportation. Integrated work plan. Version 1.0.: Washington D.C.: FAA..

Martens, M. (2007). *The failure to act on important information: where do things go wron*g. Dissertation: Free University of Netherlands: Amsterdam: The Netherlands.

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why pilots miss the green box: How display context undermines attention capture. *The International Journal of Aviation Psychology,* 14, 39-52.

NTSB (2007). Attempted takeoff from wrong runway. Aircraft accident report. Washington D.C.: National Transportation Safety Board.

Rosenthal, R. (1991). *Meta-anallytic procedures for social research*. Newbury Pk, CA: Sage Press.

Rensink, R.A. (2002). Change detection. *Annual Review of Psychology, 53*, 245-277.

Sarter, N.B., Mumaw, R., & Wickens, C.D. (2007). Pilots Monitoring Strategies and Performance on Highly Automated Glass Cockpit Aircraft. *Human Factors*, *49*, 347-357.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Science, 1*(7), 261-267.

Steltzer E.M. & Wickens, C.D. (2006). Pilots strategically compensate for display enlargements in surveillance and flight control tasks*. Human Factors, 48*, 166-181.

Taleb, N (2007). *The Black Swan. The impact of the highly improbable*. NY: Random House.

Thomas, L. C., & Wickens, C. D. (2004). Eye-tracking and individual differences in off-normal event detection when flying with a synthetic vision system display. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*. Santa Monica, CA: Human Factors and Ergonomics Society.

Wickens, C.D. (2001). Keynote address: Attention to safety and the psychology of surprise. *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus, OH: Dept. of Aerospace Engineering, Applied Mechanics, and Aviation, Ohio State University.

Wickens, C.D. & Alexander, A. L. (2009). Attentional tunneling and task management in synthetic vision displays. *International Journal of Aviation Psychology, 19*.

Wickens, C.D., Thomas, L.C., & Young, R. (2000). Frames of reference for display of battlefield terrain and enemy information: Task-display dependencies and viewpoint interaction use. *Human Factors, 42*, 660-675.

Wickens C.D., Sebok, A., Bzostek, J., Steelman-Allen, K., McCarley, J. & Sarter, N. (2009). NT-SEEV: A model of attention capture and noticing on the Flight Deck. *Proceedings of the 15th International Symposium of Aviation Psychology*. Dayton, OH: Wright State University.