

# Introduction

Most airline accidents are attributed to errors made by the flight crew (Boeing, 2004). But what does this mean? Why do highly skilled professional pilots make errors, with consequences sometimes fatal to themselves as well as to their passengers? How should we think of the role of these errors in accidents when we seek to prevent future accidents? This book seeks to address these questions in the light of what scientists are learning about the nature of skilled performance of humans performing complex tasks.

The level of safety achieved by major airlines in the United States and in most other developed countries is one of the great success stories of modern industry. The probability of becoming a fatality on a US airline flight is roughly eight in a hundred million flight segments (Sivak and Flannagan, 2003). In comparison, if one were to drive the distance of an average airline segment in a passenger car, the risk of death would be about 65 times greater (*ibid.*). This is a dramatic improvement over the early days of airline operations, when accidents were common and every airline pilot could list colleagues killed on the job (Gann, 1961; Hopkins, 1982). This improvement came about through steady advances in the design and reliability of equipment systems, operating procedures, and training over the second half of the twentieth century.

In spite of these advances the industry must constantly strive to maintain and improve safety. Every accident is a profound tragedy for the victims and their families. The horrible nature of a crash of a passenger-carrying airplane is large in the minds of the public, who rightfully demand the highest level of safety achievable. Further, the level of safety that has been achieved requires unremitting effort to maintain, in part because it is so high. The inherent nature of complex sociotechnical systems is to devolve if not constantly attended (Reason, 1997, p. 6). Economic forces contribute to this tendency. Since de-regulation in 1978 the airline industry has operated on razor-thin profit margins, and since the 11 September 2001 terrorist attacks most US companies have had to institute substantial cost-reduction programs to survive. Although no airline we are familiar with would knowingly compromise safety to cut costs, it is extremely difficult to know *a priori* whether changes designed to improve efficiency of operations and training will affect safety. Thus it is now more crucial than ever before to understand what makes the aviation system vulnerable to failure. Because most aviation accidents have been attributed historically to deficiencies in the performance of the flight crew, it is especially important to understand what makes pilots vulnerable to error.

In this book we review the 19 major<sup>1</sup> accidents in US air carrier operations from 1991 through 2000 in which crew errors played a central role, according to the National Transportation Safety Board (NTSB), the US government agency

responsible for investigating and determining the causes of all US civil aviation accidents. It is noteworthy that these 19 accidents comprise slightly more than half of the 37 major accident investigations that the NTSB conducted during this period.

Our sources of information about what happened in these accidents are the NTSB reports and associated documents that are publicly available. The NTSB conducts extensive investigations of major accidents. Analyzing wreckage, aircraft design and performance characteristics, and data from flight data recorders (FDRs) and cockpit voice recorders (CVRs), investigators are in most cases able to reconstruct the events of the accident flight to a remarkable degree.<sup>2</sup> The NTSB's investigative approach is to assign a multi-disciplinary team (the "go-team") to a major accident; the team is assisted by technically competent specialists from the aircraft manufacturer, airline, unions, and the Federal Aviation Administration (FAA), as appropriate to the circumstances of the accident. The NTSB also routinely evaluates how the crewmembers were trained, how they performed in training and in flight check rides, and how they were regarded by their fellow pilots. And of course, if the crewmembers survive, the investigators interview them extensively. Major NTSB investigations carefully scrutinize human performance issues and any crew errors and analyze how these issues and errors may have contributed to the accident. The final report<sup>3</sup> on each accident analyzes the crew's performance from both an operational and human factors perspective.

Modern airline operations are highly scripted. Airlines write flight operations manuals (FOMs)<sup>4</sup> that provide crews with instructions for all aspects of operation, both normal and non-normal (for example, emergencies). Hundreds of procedural steps are required to set up and check the correct operation of a large airplane's equipment systems before flight. These procedural steps, as well as the procedures for navigating and controlling the airplane in the air, are performed according to written scripts detailed in the FOM. These written scripts establish the correct way to perform procedures and provide standardization across pilots. Standardization is crucial because airline captains and first officers are frequently re-paired (most commonly after each three- or four-day trip), and it is common for pilots starting a trip together to be meeting for the first time if they fly for a large airline that employs thousands of pilots. Standardization also helps individual pilots to learn procedures and, with practice, remember how to execute those procedures without excessive mental demand.

Accident investigators identify crew errors in part by comparing the actions of the accident crews to the FOM written scripts and to the training provided by the airline.<sup>5</sup> Investigators also analyze the adequacy and appropriateness of the guidance contained in the FOM. The NTSB (1994a) has cited crew procedural errors as the largest category of primary errors in airline accidents (see also Karwal, Verkaik, and Jansen, 2000; Helmreich, Klinec, and Merritt, 2004); however, to interpret this observation meaningfully one must determine why deviations from procedures occur. Our analysis in this book suggests that many factors are at play. It is important to recognize that the scripts provided by FOMs represent an ideal. Actual line operations present complex situations not fully provided for by FOMs

(Loukopoulos, Dismukes, and Barshi, 2003, 2006), and the norms of actual line operations sometimes diverge from the ideal because of sociocultural, professional, and organizational factors (Reason, 1997; Helmreich and Merritt, 1998).

We use the NTSB's operational and human performance analysis of the accident as the starting point for our own analysis. We do not second-guess the NTSB as to what errors were made or how those errors contributed to the accident. However, our analysis differs in a subtle but crucial way from that of NTSB investigators because we are asking a different kind of question from the ones the Board is charged to address. The NTSB's charter requires it to determine the probable cause of each specific accident. As far as possible, the human performance group of the investigation team attempts to identify factors that may have contributed to the crew's errors; however, to cite such factors the NTSB must be reasonably certain that these factors did in fact contribute significantly to the errors and, by extension, to the particular accident under consideration.

Unfortunately, in many cases this is not possible. For example, consider the case of a crew preparing for takeoff who inadvertently fail to set the flaps to the takeoff position. (Setting the flaps is a highly practised procedural step that the crew would have performed thousands of times before without failure.) Although the NTSB might be able to determine with confidence that the crew failed to set the flaps, it is usually not possible to determine with any certainty why the crew overlooked a step. Indeed, if the crew survives and are interviewed, they themselves typically cannot explain the oversight. The investigators may be able to identify various factors, such as distraction or a poorly designed checklist that might contribute to an error such as this, but rarely can the degree of influence of such factors be determined. The most common result in a situation like this is that the NTSB will identify the error made by the crew (failure to set the flaps), discuss the human performance issues that could have affected the crew's performance, but stop short of drawing conclusions that link the error to the underlying causes.<sup>6</sup>

We face the same limitation as NTSB investigators and do not attempt to go beyond their reports in analyzing what actually caused the accident crews to make the errors they did. What we do is ask a subtly different question, one that is profoundly important to aviation safety: if a population of pilots with experience and skills comparable to those of the accident crew faced a situation similar, though not necessarily identical, to that which confronted the accident crew, would this population of pilots be vulnerable to making the kinds of errors made by the accident crew and, if so, why? Thus our purpose is different from that of the NTSB, which must attempt to determine what actually happened to the specific crew in a specific accident, to the extent it is possible to do so. In contrast, we attempt to understand the nature of vulnerability of all skilled pilots to error in situations similar to those faced by the accident pilots.

It would be impossibly cumbersome for us and tedious to the reader if, in discussing the many errors identified in the accident reports, we were to use over and over the formulation "A population of pilots similar to the accident crew might have had some degree of vulnerability to making this particular error made by one

of the accident pilots because ...” Thus for simplicity we typically say “The captain (or first officer) might have been vulnerable to this error because ...” but periodically remind the reader that we are really talking about a population. Behavioral scientists often use the percentage of a population of individuals making an error in a situation as equivalent to the probability that an individual from that population would make that error.

Drawing upon rich literatures in experimental psychology and human factors, especially research on cognitive mechanisms underlying the skilled performance of experts, we are able to identify factors that make all pilots vulnerable to specific kinds of errors in various situations. A central perspective emerging from the scientific literature is that the occasional errors made by pilots and other skilled experts occur in a somewhat random fashion; thus scientists speak of factors *influencing the probability of errors* rather than *causing errors*. Multiple factors, not all of which can be determined and measured, interact to produce an error in a given instance.

Many people assume that if an expert in some domain (aviation, medicine, or any other) makes an error, this is evidence of lack of skill, vigilance, or conscientiousness. This assumption is both simplistic and wrong (see “bad apple theory” in Dekker, 2002). Skill, vigilance, and conscientiousness are of course essential for safe, effective performance but are not sufficient. A particularly problematic misconception about the nature of skilled human performance is that, if experts can normally perform some task without difficulty, then they should always be able to perform that task correctly. But in fact experts in all domains from time to time make inadvertent errors at tasks they normally perform without difficulty. This is the consequence of the interaction of subtle variations in task demands, incomplete information available to the expert performing the task, and the inherent nature of the cognitive processes that enable skilled performance.

In this book we discuss at length the vulnerabilities of human cognitive processes such as attention, memory, and decision-making. Those vulnerabilities must be considered in the appropriate context. Computer technology has advanced to the point that it is in principle possible to operate an aircraft from takeoff to landing without human intervention. However, for compelling reasons, this is not done. Computers have extremely limited capability for dealing with unexpected and novel situations, for interpreting ambiguous and sometimes conflicting information, and for making appropriate value judgments in the face of competing goals. These functions are appropriately reserved for human experts. Humans are able to perform some tasks far beyond the capabilities of computers because of the way our brains have evolved, but part and parcel of our unique abilities is inherent vulnerability to characteristic forms of error in certain situations. As will become apparent in our analysis of airline accidents, human skill and vulnerability to error are closely linked through underlying cognitive processes.

Human operators are expected to make up for the deficiencies in the design of systems, and this is manifest in aviation operations. Airline crews routinely deal with equipment displays imperfectly matched to human information-processing characteristics, respond to system failures, and decide how to deal with threats

ranging from unexpected weather conditions to passenger medical emergencies. Crews are able to manage the vast majority of these occasions so skillfully that what could have become a disaster is no more than a minor perturbation in the smooth flow of high-volume operations. But on the rare occasions when crews fail to manage these situations, it is detrimental to the cause of aviation safety to assume that the failure stems from deficiency of the crews. Rather, these failures occur because crews are expected to perform tasks at which perfect reliability is not possible for either humans or machines. If we insist on thinking of accidents in terms of deficiency, that deficiency must be attributed to the overall system in which crews operate.

Contributing to the misunderstanding of the vulnerability of experts to error is that the presence and interaction of factors contributing to error is probabilistic rather than deterministic. Accidents are rarely, if ever, caused by a single factor but rather by a complex interplay of multiple factors, combining in ways driven in large degree by chance, each factor influencing the effects of the others (Reason, 1997). After analyzing many accident reports – not just those described in this book – we strongly suspect that small random variations in the presence and timing of these factors substantially affect the probability of pilots making errors leading to an accident. In addition to variations in the interaction of external factors, we must recognize that individuals themselves vary moment to moment in the fine-grained detail of their responses to external factors. Thus if it were somehow possible to identically replicate in large numbers the crew of an accident and put each replica crew in exactly the same initial situation, they would not all perform identically, and the differences would grow as the situation unfolded and their reactions altered the interplay of external factors.

To protect and improve aviation safety we must understand what makes pilots vulnerable to error and must understand the interplay of factors contributing to that vulnerability. This book is an attempt to shed light on the nature of the vulnerability of pilots and other experts to error when performing tasks at which they are highly skilled. Fortunately scientists have now learned a fair amount about the cognitive processes and the conditions that shape the form and probabilities of errors. To a large degree the errors made by experts are driven by four factors:

- 1) specific characteristics of the tasks performed;
- 2) events in the environment in which tasks are performed;
- 3) demands placed on human cognitive processes by task characteristics and environmental events; and
- 4) social and organizational factors that influence how a representative sample of experts would typically operate in particular situations (including both the ideals expressed in training and formal guidance and the ‘norms’ for actual operations).

Our analysis of pilots’ vulnerability to error systematically considers the interaction of these four types of factors in the kinds of situation faced by the pilots in the 19 accidents examined in this book. We also adhere to the assumption of local

rationality (Woods and Cook, 1999): experts typically do what seems reasonable to them at the time, given their understanding of the circumstances. This does not mean that experts never fail to be conscientious, but we suggest that errors are not *de facto* evidence of lack of conscientiousness and that the burden of proof falls on anyone claiming an expert was not conscientious.

In seeking to understand the errors made by expert pilots, it is crucial to avoid hindsight bias (Fischhoff, 2003; Dekker, 2002, pp. 16–20), a term cognitive scientists use to describe distortion of evaluators' judgments by knowledge of the outcome of the situation that is being evaluated. Knowing the disastrous outcome of a flight makes it easy to identify things the crew could have done differently to prevent the accident. But of course accident crews cannot foresee the outcome of their flights; as far as they can tell, up until the moment when things start to go wrong, they are conducting flights as routine as thousands they have flown before.

We believe our approach can contribute to aviation safety by helping the understanding of the nature of vulnerability of skilled experts to error, which in turn will lay a foundation for developing countermeasures to reduce that vulnerability. However, there are limitations to our approach: it is inherently speculative. Accident reports do not provide sufficient fine-grain detail of the many factors involved to determine the probability of a particular kind of error, and scientific understanding of the cognitive processes underlying skilled human performance is far from complete. Thus we cannot say what percentage of a large population of airline pilots would have made errors similar to those made by the accident crew if placed in their exact situation, but we are reasonably certain that some of this large population would make similar errors – however, this is not an indication of deficiency on the part of those pilots who did make errors. The probability of any pilots – including the accident pilots – making a particular error in a given situation is quite low, and the probability of that error combining with other circumstances to create an accident is extraordinarily low. But because exposure to opportunity to error is very high – almost 17 million flights are conducted worldwide every year (2003 data, commercial jet airplanes, >60,000 pounds maximum gross weight) (Boeing, 2004) and each flight has numerous opportunities for error – errors do occur regularly. We can say with reasonable confidence from our analysis of cognitive vulnerabilities in the 19 accidents discussed in this book, and from the fact that the kinds of error in these accidents also occur regularly in non-accident flights (Loukopoulos et al., 2003, 2006), that few if any pilots are immune from making the kinds of error that occurred in these accidents. Thus it is essential to understand the true nature of vulnerability to error in order to reduce that vulnerability, to devise ways to help pilots catch errors before they become consequential, and to making the aviation system resilient to errors that are not caught.

To provide a good sample of the range of situations and the range of errors leading to accidents, we chose to treat all 19 major US accidents attributed at least partly to crew error in the period 1991 through 2000. This allowed us to identify common themes underlying the accidents; these common themes and their implications are discussed in the final chapter (Chapter 21). This approach also allows our book to

serve as a companion piece to a previous study by the NTSB, for which the second author served as the principal investigator while on the NTSB staff: *A Review of Flightcrew-involved Major Accidents of US Air Carriers, 1978 through 1990* (NTSB, 1994a). Accidents in the NTSB study were selected using criteria similar to our own. The NTSB study provided a statistical analysis of characteristics of the operating environments, characteristics of crewmembers, errors committed, and the contexts in which the errors occurred in the 37 accidents during that 12-year period. Among the interesting findings of this study were that the captain was flying in more than 80 per cent of the accidents, and that the crew were flying together for the first time on the day of the accident in 73 per cent of accidents for which data were available. Chapter 20 of our book provides a comparable statistical analysis that updates and extends the NTSB study.

The 19 accident chapters of this book follow a common format. We first provide a brief description of the accident and the NTSB conclusions about the causes of the accident and contributing factors. We then discuss the significant events leading to the accident in chronological order, first briefly describing the event and then discussing crew actions, errors identified by the NTSB, and our thoughts on why pilots may be vulnerable to these kinds of errors. Each accident chapter concludes with a short summary and discussion of implications of our analysis. The accident chapters can be read independently of each other, so readers who are especially interested in a particular accident or particular type of situation can go directly to the chapter of interest. The statistical summary chapter comes after these accident chapters, and in the final chapter we attempt to draw together what we have learned in this study and suggest specific ways to improve aviation safety.

This book is written for a diverse audience. It will of course be of particular interest to the aviation operational community and to scientists studying aircrew performance. Beyond that it should be useful for anyone interested in aviation safety and those interested in understanding why skilled experts in any domain make errors, sometimes with disastrous consequences. To make the book as widely accessible as possible we explain technical terms and concepts in endnotes to the extent practical. For further information the reader without experience in aviation might consult books such as *The Turbine Pilot's Flight Manual* (Brown and Holt, 2001). Also, Walters and Sumwalt (2000) provide a very readable summary of a wide cross section of aviation accidents.

## Notes

- 1 The NTSB defines an accident as major when a 14 CFR 121 aircraft was destroyed or multiple fatalities occurred or one fatality occurred and a 14 CFR 121 aircraft was substantially damaged. However, the NTSB adopted this definition recently and thus did not use these criteria in selecting the cases for its 1994 safety study on flight crew performance in major US air carrier accidents (NTSB, 1994a). To facilitate comparison with the earlier study, we selected these 19 accidents using two criteria that closely matched the ones used by the NTSB in 1994: (1) the NTSB conducted a major investigation (signified by an

- NTSB accident ID in the “MA” series) or produced a major accident report in the NTSB AAR report series, and (2) the NTSB cited deficient performance of the flight crew as a causal or contributing factor.
- 2 CVRs and FDRs (see glossary) are essential tools for accident investigation. In this book we make extensive use of the NTSB transcripts of CVR recordings (the actual recordings are not publicly available) and of FDR data in NTSB reports. The CVR data are our primary factual source for what the crewmembers were doing through the sequence of events in an accident; however, this source of information is inherently incomplete. The recording captures only utterances and other audible sounds and provides no information about non-verbal communications or about movements of the crew such as pointing and touching. Further, transcribing the recording inherently requires varying degrees of interpretation. The FDR data, used in conjunction with the CVR transcript, help flesh out the picture of crew actions, but it is not possible to know everything the crew was thinking and doing.
  - 3 Most major aviation accident investigations performed by the NTSB result in a detailed final report known as a “Bluecover” report (the name is derived from an earlier – but not current – color scheme) that follows the ICAO Appendix 13 format for an accident report, first summarizing relevant factual information about the accident, then analyzing the facts, and finally reporting the agency’s findings, causal determinations, and recommendations for safety improvement. Such a report was available for 17 of the 19 accident cases that we review in this book. Sometimes the NTSB publishes the results of a major investigation in a less extensive, “summary” or “brief format” report. In most of these summary reports the discussion of the facts and analysis of the accident is much less extensive than in a Bluecover report. For the two major accident cases we analyze for which the NTSB did not produce a major accident report, we also reviewed publicly available background reports by NTSB investigators (“group chairman factual reports”) in the areas of flight operations, human performance, and data recorders.
  - 4 The FOM is based on but more detailed than operating manuals provided by the aircraft manufacturer.
  - 5 And from the standpoint of scientists studying skilled performance of experts, airline operations provide an excellent domain for investigation because it is easier to distinguish correct from incorrect performance than in many other domains of skilled performance.
  - 6 The NTSB’s ultimate goal is to identify safety improvements that can prevent recurrence of accidents from the same causes. Recognizing the need to identify factors that contribute to the errors of aviation personnel, the NTSB increasingly discusses factors that might have contributed to errors and makes related safety recommendations, even if it is not sufficiently certain of the degree of influence of those factors to cite them as probable causes or contributing factors in a particular accident.