University of
Southern
Queensland

MAINTENANCE CHECK FLIGHT ACCIDENTS

(A NEW APPROACH TO AIR SAFETY INVESTIGATIONS)

A Thesis submitted by:

Zsolt Jay Nagy (MSc)

For the award of

Doctor of Philosophy

2023

ABSTRACT

In the wake of several serious post-maintenance check flight accidents and incidents, the airline industry adopted a new operational risk management framework. This research study found a misalignment between the traditional safety response and the underlying causal patterns elicited from accident investigation reports. From a system safety perspective, the industry must recognise the novel safety hazards associated with highly automated airliners, otherwise, the unmitigated systemic risk will lead to further accidents. Owing to the lack of studies and safety statistics on the topic, the research project started with building a comprehensive catalogue of non-routine flight occurrences for Western-built commercial jet airliners. Once maintenance check flight investigation records were identified, systemic causal and contributing factors were extracted and assembled into a knowledge map. Finally, grounded in an embedded multi-case study across three different airliner generations, a refined accident model was developed for check flight accident causation. Common patterns were traced to the underlying regulatory framework and the way systemic causes propagate through operational and design pathways. The study concluded with practical recommendations for the industry, supported by further research ideas which may offer additional insights into the non-routine safety problem.

CERTIFICATION OF THESIS

I, Zsolt Jay Nagy, declare that the PhD Thesis entitled *Maintenance Check Flight Accidents (A New Approach to Air Safety Investigations)* is not more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, and references. The thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, the thesis is my own work.

Date: 12 Dec 2023

Endorsed by:

Associate Professor Tarryn Kille
Principal Supervisor

Dr. Wayne Martin
Associate Supervisor

Student and supervisor signatures of endorsement are held at the University.

ACKNOWLEDGEMENTS

First, let me acknowledge my supervisory team:

Dr. Wayne Martin carefully guided this project from day one. I could not ask for a more supportive PhD supervisor. Thanks to your sage advice and patience over the years, I finally made it to the finish line.

Professor Emeritus Paul Bates and Honorary Professor Pat Murray. I am fortunate to have been supported by your wisdom when formulating the research problem, documenting initial results, and drafting my first academic papers.

Dr. Tarryn Kille, my principal supervisor at UniSQ Aviation for the final year of the research project, and Dr. Maneerat Tianchai. I am grateful for your review of the final draft and your guidance in meeting the formal degree requirements.

Dr. Constantin Ferroff was always available to review my manuscripts, validating the research framework and greatly improving my academic writing.

I also need to mention my gratitude for the support received from some very special people:

To Helge, for your enduring friendship and your timely reminders about what is the most important in life.

To our friends in France, for adopting me when the international borders were slammed shut and I could not return home to my family.

Finally, a special thank you to my industry network - including the many airline and test pilots, maintenance engineers, system architects, design engineers, and safety investigators - for sharing their insights and concerns with me. I hope that this study will contribute to safer non-routine flight operations.

This research has been supported by the Australian Government Research Training Program Scholarship.

DEDICATION

This research project is dedicated to the memory of Janos Nagy Jr. and Prof. Istvan Steiger. Although my father and my eminent Aeronautics Professor were the inspiration to pursue my doctoral degree, they were unable to see my graduation.

The thesis is dedicated to my family: to my wife Noemi, whose love and support give me strength, no matter what challenges we may face in life. And to Daniel, Nikol, and Natalie, for always believing in your Dad. I can’t wait to spend more time with you all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AAIB	Air Accidents Investigation Branch
ABX	Airborne Express
ADG	Air Driven Generator
ADM	Aeronautical Decision Making
AMM	Aircraft Maintenance Manual
ANZ	Air New Zealand
AOA	Angle of Attack
AOB	Angle of Bank
ASN	Aviation Safety Network
ATC	Air Traffic Control
ATSB	Australian Transport Safety Bureau
AvH	Aviation Herald
BEA	Bureau d’Enquetes et d’Analyses
BFU	Bundesstelle fur Flugunfalluntersuchung
CAA	Civil Aviation Authority
CAM	Customer Acceptance Manual

CAST	Commercial Aviation Safety Team
CDFS	Customer Demonstration Flight Schedule
CFIT	Controlled Flight Into or Toward Terrain
CFR	Code of Federal Regulations
CIAIAC	Civil Aviation Accident and Incident Investigation Commission
CRM	Crew Resource Management
CSE	Cognitive Systems Engineering
CVR	Cockpit Voice Recorder
EASA	European Aviation Safety Agency
FAA	Federal Aviation Administration
FBW	Fly-by-Wire
FCM	Fuzzy Cognitive Map
FDR	Flight Data Recorder
FMS	Flight Management System
FRAM	Functional Resonance Accident Model
FSF	Flight Safety Foundation

FSS	Flight Safety Systems
GPWS	Ground Proximity Warning System
GXL	Excel Airways Germany
HF	Human Factors
HFACS	Human Factors Analysis and Classification System
HRO	High Reliability Organization
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
LOC-I	Loss of Control – In Flight
MAC	Mid-Air Collision
MCF	Maintenance Check Flight
MRO	Maintenance Repair and Overhaul
MTOW	Maximum Take-off Weight
NASA	National Aeronautics and Space Administration
NDM	Naturalistic Decision Making
NPA	Notice of Proposed Amendment
NRC	Nuclear Regulatory Commission

NRFO	Non-Routine Flight Operation
NTSB	National Transportation Safety Board
OEM	Original Equipment Manufacturer
PF	Pilot Flying
PIC	Pilot-In-Command
PM	Pilot Monitoring
RCA	Root Cause Analysis
RPD	Recognition-Primed Decision
RWY	Runway
SARPs	Standards and Recommended Practices
SCF	System Component Failure or Malfunction
SHEL(L)	Software, Hardware, Environment, Liveware
SOP	Standard Operating Procedures
STAMP	Systems-Theoretic Accident Model and Processes
STC	Supplemental Type Certificate
TAIC	Transport Accident Investigation Commission
TAWS	Terrain Awareness and Warning Systems

TEM	Threat and Error Management
TSBC	Transportation Safety Board of Canada
WBA	Why-Because Analysis

CHAPTER 1 - INTRODUCTION

1.1 Background

The airline industry is justifiably proud of its record in achieving and maintaining ultra-safe levels during routine commercial operations where the annual fatal accident rate is consistently less than 1 per a million departures (Airbus, 2022; Amalberti, 2001; Boeing, 2022). Airline pilots are trained to conduct routine line operations in a safe manner and in line with standard operating procedures (SOP). Airline operators must ensure that their pilots follow rules and regulations in every phase of flight. Unbeknown to the public, non-routine flights contribute to that excellent safety record by shifting higher-risk operations to non-revenue sectors, when only essential crew members are on board.

Non-routine flying is a broad operational category that includes all flight operations other than scheduled or chartered revenue flights. For mainline and regional airlines, common examples include positioning and ferry flights, training flights, airshows (low altitude demonstrations), and functional and operational check flights, following maintenance action or modification programs (Federal Aviation Administration [FAA], 2008a, 2016). According to FAA safety statistics for Part 121 airline operations over ten years between 1998 and 2007, approximately 25 percent of accidents involving turbine powered airplanes have occurred during non-routine flight operations (NRFO) (FAA, 2008b).

During recent decades several high-profile accidents and serious incidents highlighted the elevated risk profile of non-revenue/non-routine airline operations and experimental flight test certification programs:

Jun 1988 – A320, Mulhouse-Habsheim, France: The accident involved an A320, the first digital fly-by-wire (FBW) airplane type certified for commercial air transport. Air France operated the demonstration flight at a local airshow, soon after introducing the A320 to its fleet. The airplane flew over the runway at a low height, with engines at flight idle and the angle of attack (AOA) increasing. The airplane sank into the forest a short distance beyond the airfield, impacted the trees, and caught fire on the ground. Three passengers lost their lives. The investigation concluded that flight crew error, especially the late application of go-around power, was the probable cause of the accident (Bureau d’Enquetes et d’Analyses [BEA], 1989).

Jun 1994 – A330, Toulouse, France: At the time of the accident flight, the A330 was undergoing a flight test in preparation for certifying the autopilot for Category III operations with Pratt and Whitney engines. The flight was supposed to test the autopilot’s speed reference system (SRS mode), which was meant to control the aircraft’s speed and angle of attack in case of an engine out go-around condition. Three crew members, including the Airbus chief test pilot, an airline training captain (pilot flying), an Airbus flight test engineer, and four observers were fatally injured. The airline captain was not rated for experimental flight test operations (Direction general de l'armement [DGA], 1994).

Dec 1996 – DC-8-63, Narrows, Virginia, United States: A DC-8-63 freighter was destroyed during a post-modification evaluation flight, fatally injuring all three crew members and three observers on board. The National Transportation Safety Board (NTSB) determined that the probable causes of the accident were: pilot error, and the failure of the airline to establish a formal check flight program (NTSB, 1997).

Oct 2004 – CRJ-200, Jefferson City, Missouri, United States: The airplane was on a positioning flight from Little Rock to Minneapolis. During the flight, both engines flamed out after a pilot-induced high altitude aerodynamic stall. The engines could not be restarted due to a core lock condition. The captain and the first officer were killed, and the airplane was destroyed (NTSB, 2007).

Nov 2008 – A320, Perpignan, France: The airplane operated a post-maintenance check flight, in the context of ending a lease agreement, when it was destroyed upon impact off the coast of Canet-Plage. Investigating this major accident involved two airlines and several safety boards and aviation authorities from around the world. The investigation concluded that the principal cause of the accident was an improvised test of flight envelope protections, while blocked angle of attack sensors prevented those protections from triggering. The Perpignan tragedy created significant interest and concern within the airline industry and the final investigation report played a key role in shaping the current safety response (BEA, 2010).

Jan 2009 – B737-700, Norwich, United Kingdom: The airplane operated a combined check flight and customer demonstration program, having just completed a maintenance visit, when it experienced a serious in-flight upset and loss of control incident. The airplane violently pitched down and lost approximately 9,000 feet before the pilot was able to recover and safely land (Air Accidents Investigation Branch [AAIB], 2010b). This serious incident served as another major wake-up call to the industry, confirming that the check flight safety problem is not limited to any manufacturer or design philosophy.

Nov 2009 – Falcon 2000, Biggin Hill, United Kingdom: The airplane suffered substantial damage during a post-maintenance operational check flight, involving a series of high-speed taxi tests. During the 8th taxi run the crew brought the aircraft to a stop, when the control tower informed them that there was a fire under the wing. The crew and passengers safely abandoned the airplane (AAIB, 2010a).

Apr 2011 – G650, Roswell, United States: The accident involved an experimental Gulfstream GVI (G650) airplane, registered to, and operated by Gulfstream, as part of its G650 flight test program at Roswell, New Mexico. The accident occurred during a planned one-engine-inoperative (OEI) takeoff when a stall on the right outboard wing produced a rolling moment that the flight crew was not able to control. The two pilots and the two flight test engineers were fatally injured, and the airplane was substantially damaged by impact forces and a post-crash fire (NTSB, 2012).

May 2011 – Falcon 7X, Kuala-Lumpur, Malaysia: A Dassault Falcon 7X suffered an in-flight upset (pitch trim runaway in normal law) during descent to Kuala Lumpur Airport. The airplane was operating a positioning flight from Germany, registered to a Swiss operator. The three crew members were not injured. The investigation led to a range of safety recommendations, including new system safety assessments for electronic equipment and software, and improved training for taking over control of airplanes equipped with non-coupled control sticks (BEA, 2016).

Nov 2018 – E190, Alverca, Portugal: An Embraer E190 suffered a major in-flight upset and loss of control during a combined post-maintenance validation and ferry flight. Immediately after takeoff, the crew declared an emergency, having completely lost control of the airplane. The crew managed to safely land the airplane on the third attempt, aided by air traffic control (ATC) and the Portuguese Air Force. One crew member received minor injuries. The aircraft was overstressed during the event and was later written off. The investigation found that the aileron control system cables were incorrectly installed during the maintenance event (Gabinete de Prevenção e Investigação de Acidentes com Aeronaves e de Acidentes Ferroviários [GPIAAF], 2020).

Apr 2019 – Global 5000, Berlin, Germany: A Bombardier Global 5000 from the Federal German Government's fleet was on a ferry flight to Cologne after a routine maintenance event. Shortly after takeoff, the crew experienced problems with the flight controls and decided to return to Berlin-Schonefeld. During a normal turn, the aircraft departed controlled flight. The crew recovered the airplane, but experienced severe control difficulties during landing, leading to both wings contacting the ground before the airplane settled on the main gears. The crew was not injured, but the airplane was severely damaged (Directorate of Aviation Safety Bundeswehr, 2021).

As illustrated by the above examples, the problem is not limited to any airplane type, airliner generation, or any specific type of non-routine operation. Within the non-routine category, functional and operational check flights remain an important element of an airline's continuing airworthiness management responsibilities. Despite the elevated risk profile, the actual magnitude of operational safety risk is not well understood by the airline industry and the safety research community.

This research project aims to highlight and address that knowledge gap in the context of post-maintenance check flight operations.

The thesis explores the differences between routine and non-routine flying, reviews the new safety challenges originating from new airliner designs, and provides an overview of airline accident investigations and new models for complex system accidents. Furthermore, the thesis intends to serve as a foundation for further research, by revisiting and critically evaluating landmark accidents, and proposing a new approach to air safety investigations.

1.2 Statement of the Problem

The airline industry adopted a *plan and prepare* approach in response to a series of high-profile post-maintenance test flight events (Airbus, 2015; Flight Safety Foundation [FSF], 2011a, 2011b). Better planning and preparation are proven risk mitigation measures for scenarios foreseen in advance, and they may be credited for achieving some operational safety gains (Roland & Moriarty, 1990). At the same time, it is not clear how this approach would have prevented past check flight accidents or serious incidents. The current safety management framework fails to account for a fundamental design assumption: even in the latest generation of commercial jet airplanes, the human pilot remains the last line of defence whenever automated systems are unable to cope with unforeseen or unexpected scenarios (Dismukes et al. 2007). Better planning and preparation are not suitable risk controls for those novel challenges.

Maintaining ultra-safe routine commercial operations is a priority for the airline industry and transport safety agencies (Amalberti, 2001). Despite the elevated risk profile associated with non-routine flying, the actual magnitude of operational safety risk is not understood by the industry and the safety research community (The International Federation of Air Line Pilots' Associations [IFALPA], 2019; FAA, 2016; FSF, 2011a; Poprawa, 2015). Although accident investigation processes have substantially improved over the years, most accident causation models still focus on elusive root causes (Leveson, 2011). Blaming technical pilots for making intuitive decisions, which in hindsight may appear incorrect to an external observer, is a counter-productive approach (Kahneman & Klein, 2009). If we do not know what contributes to the safety problem, it is very difficult to introduce the necessary systemic changes to non-routine operations.

As more reliable commercial jet airplanes enter the worldwide fleet, mechanical hardware failures will continue to decrease. At the same time, it remains virtually impossible to tolerate all design faults in complex avionics systems (Hitt & Mulcare, 2000). It follows that an airline pilot cannot be fully prepared for unanticipated errors generated by unknown or unpredictable design faults. Next generation aircraft are likely to adopt a system architecture with a higher level of automation and allocate more decision making authority to automated systems (Abbott et al., 1996; Parasuraman et al., 2000). These new system architectures must be carefully evaluated to ensure that future airplane models will not experience even worse safety outcomes during non-routine flying.

1.3 Purpose of the Study

The purpose of this research study is to develop causal explanations for maintenance test flight accidents. Causal factors and timelines annotated in air safety investigation reports provide empirical evidence for building an initial knowledge map about primary categories and their causal relationships.

A second purpose is to explore the relationship between the safety response implemented by the industry and the common patterns and interwoven themes identified across multiple critical incidents. The deeper causal structure revealed by a comparative multiple case study serves as a new theoretical framework for validating existing and additional risk controls and safety recommendations.

1.4 Significance of the Study

This thesis contributes valuable insight and a new theory into the development and implementation of a suitable airline industry response to the critical safety challenge posed by post-maintenance test flights.

Despite the considerable interest and concerns expressed by the check flying community, there is very little academic research on non-routine flight operations. Poprawa’s (2015) contribution is a notable exception, who found that maintenance test flights remain at least one or two magnitudes more risky than normal commercial sectors. The research study creates new knowledge by introducing a revised theoretical framework for accident investigations and has practical application in offering important safety recommendations for the airline industry.

1.5 Research Questions

The overarching objective of this study is to discover under what conditions, and through what causal paths, maintenance test flight accidents unfold. The three specific research questions addressed are:

Research Question 1

What causes and contributing factors are annotated in relevant check flight investigation reports?

Research Question 2

Are there any common patterns in check flight occurrences? If there are common patterns, what do they reveal about the deeper structure of check flight accidents?

Research Question 3

Is there an alignment between underlying patterns and the risk control framework implemented by the airline industry?

1.6 Delimitations

Lunenburg and Irby (2008, p. 134) define delimitations as "self-imposed boundaries set by the researcher on the purpose and scope of the study". In this study, the primary delimitation is reflected in the dissertation title: from the broader non-routine flight operation class, only maintenance check flight occurrences are selected for detailed analysis. This relatively narrow scope is necessary to ensure that all cases belong to the same subclass, including closely aligned mission objectives and similar supporting organisational elements.

As a second delimitation, only Western-built commercial jet airplanes in mainline service are included in the study. Eastern-built models are excluded, as no reliable safety investigation reports and statistics are available. At the same time, the decision to limit the scope to mainline commercial jets improves the reliability of the study, as the number of non-routine accidents and incidents uncovered in the data collection phase can be compared with Boeing's safety statistics (Boeing, 2022). Boeing's (2022) annual statistical summary is the only published literature which contains some basic information about non-routine accidents and hull losses suffered by Western-built airliners. Furthermore, Boeing's (2022) source data is limited to a defined set of commercial jets, excluding turboprop airliners, and as such, turboprop models are excluded from the study.

1.7 Limitations and Assumptions

The following limitations arose from the selected research methodology, available data, and method of data analysis:

1. The sample of check flight occurrences is drawn from final investigation reports published by transport safety agencies, therefore limiting the sample size.
2. The author’s position as an active airline system safety practitioner may be perceived as a limitation on the validity of the study. This limitation is addressed by adapting the qualitative research design as outlined in this section.
3. Production test flight occurrences were removed from scope to address a perceived conflict of interest during the study period.

The limited sample size is the primary limitation of the study. Many factors influence the decision to proceed with investigating a non-routine safety occurrence. Some airlines prefer to handle non-routine safety events as an internal matter and minor incidents may not be notified to local authorities. Even if a non-routine incident is formally reported by the airline, resource constraints may prevent the relevant transport safety agency from fully investigating the event or progressing the investigation process beyond filing a brief factual account.

Traditionally, the researcher’s background and identity have been considered as bias, especially when the research is closely related to the field of study. Strauss (1987) highlighted the value of incorporating the researcher’s own experience as experiential data, a position which later gained considerable theoretical support (Denzin & Lincoln, 2011). While Maxwell (2009) underlined the importance of gaining insights and a credible source of validity checks by being immersed in the field, the author’s position relative to the thesis scope needs to be explicitly stated.

The author is a system safety practitioner who is active in the airline safety community and is part of the target population. The author’s position may raise the question of subjectivity, due to the difficulty involved in clearly delineating opinion and expert insight when documenting qualitative research findings. The perceived limitation is addressed by employing a qualitative research design framework that clearly states whenever the author’s own experience is relied upon, including case selection, analysis, theory generation, and safety recommendations.

Over the course of working on this research project, the author was employed by a major airframe manufacturer as an airline operations expert. The former employment relationship can raise similar concerns about potential bias, and as a result, all production test flight occurrences were removed from scope. As an additional mitigation, the research design was adapted to ensure that case studies represent a balanced selection of airliner generations and design philosophies.

In addition to the above-mentioned limitations, the study incorporates the following assumptions: a) final investigation reports reflect a valid and logical construct of accident causes; b) International Civil Aviation Organization (ICAO) definitions prevail when an occurrence is classified as an accident or incident; and c) operational check flights operated as a routine commercial sector are included in the study, provided there was an open entry in the technical log that required the pilots to carry out an in-flight maintenance verification task.

1.8 Definition of Terms

The study adopts ICAO (2020) definitions when differentiating between *accidents, serious incidents, incidents*, and *occurrences*. Additionally, for the purposes of data analysis, individual terms for *accident rates* and *hull losses* are borrowed from Boeing's (2022) safety reports. Finally, the EASA definition of *maintenance check flight* is central to the study and referred to throughout the thesis.

Accident

An occurrence associated with the operation of an aircraft which, in the case of a manned aircraft, takes place between the time any person boards the aircraft with the intention of flight until such time as all such persons have disembarked, or in the case of an unmanned aircraft, takes place between the time the aircraft is ready to move with the purpose of flight until such time as it comes to rest at the end of the flight and the primary propulsion system is shut down, in which:

a) a person is fatally or seriously injured as a result of:

being in the aircraft, or direct contact with any part of the aircraft, including parts which have become detached from the aircraft, or direct exposure to jet blast, *except* when the injuries are from natural causes, self-inflicted or inflicted by other persons, or when the injuries are to stowaways hiding outside the areas normally available to the passengers and crew, or

b) the aircraft sustains damage or structural failure which:

adversely affects the structural strength, performance, or flight characteristics of the aircraft, and would normally require major repair or replacement of the affected component, *except* for engine failure or damage, when the damage is limited to a single engine (including its cowlings or accessories), to propellers, wing tips, antennas, probes, vanes, tires, brakes, wheels, fairings, panels, landing gear doors, windscreens, the aircraft skin (such as small dents or puncture holes), or for minor damages to main rotor blades, tail rotor blades, landing gear, and those resulting from hail or bird strike (including holes in the radome); or

c) the aircraft is missing or is completely inaccessible (ICAO, 2020, p. 19).

Accident Rates

In general, this expression is a measure of accidents per million departures. Departures (or flight cycles) are used as the basis for calculating rates because there is a stronger statistical correlation between accidents and departures than there is between accidents and flight hours, or between accidents and the number of airplanes in service, or between accidents and passenger miles or freight miles (Boeing, 2022, p. 20).

Hull Loss

Airplane totally destroyed or damaged and not repaired. Hull loss also includes, but is not limited to, events in which:

- The airplane is missing.
- An aircraft is considered to be missing when the official search has been terminated and the wreckage has not been located.
- The airplane is completely inaccessible (Boeing, 2022, p. 20).

Incident

> An occurrence, other than an accident, associated with the operation of an aircraft which affects or could affect the safety of operation (ICAO, 2020, p. 20).

Maintenance Check Flight

> A flight of an aircraft with an airworthiness certificate or with a permit to fly which is carried out for troubleshooting purposes or to check the functioning of one or more systems, parts or appliances after maintenance, if the functioning of the systems, parts or appliances cannot be established during ground checks… (Regulation (EU) 2019/1384, 2019, p. 4)

Occurrence

> Any accident or incident associated with the operation of an aircraft (ICAO, 2015, p. 11).

Serious incident

> An incident involving circumstances indicating that there was a high probability of an accident and associated with the operation of an aircraft which, in the case of a manned aircraft, takes place between the time any person boards the aircraft with the intention of flight until such time as all such persons have disembarked, or in the case of an unmanned aircraft, takes

> place between the time the aircraft is ready to move with the purpose of flight until such time as it comes to rest at the end of the flight and the primary propulsion system is shut down.
> (ICAO, 2020, p. 21).

1.9 Thesis Structure

The thesis is presented in nine main chapters:

This chapter introduces the safety problem and the purpose of the research project, along with the research questions, limitations, and key assumptions.

The next chapter summarizes the available literature, focusing on the differences between routine and non-routine flying, accident investigation methods and findings, and the evolving safety response, as shaped by the in-service experience with highly automated commercial jet airplanes.

The third chapter outlines the qualitative research framework applied, including an empirical phenomenological study that combines check flight investigation results into a single knowledge map, followed by a comparative multiple case study that generated a new grounded theory of maintenance check flight accident causation.

Chapter four is dedicated to sharing the results from the initial part of the study, including historical investigation records, first- and second-order constructs built from investigation reports, and the limitations of the knowledge map.

Chapters five, six, and seven are dedicated to the case studies that focus on within-case analysis of landmark maintenance test flight events. Chapter eight presents a detailed cross-case analysis and a new grounded theory that fits common patterns in check flight accidents.

Finally, the last chapter concludes the study, including a summary, practical implications, and safety recommendations on how to further mitigate the safety risks associated with maintenance check flight operations.

CHAPTER 2 – LITERATURE REVIEW

2.1 Introduction

This chapter provides the rationale for researching maintenance check flight accidents and incidents. There are very few peer-reviewed articles on this topic, therefore, the literature review had to rely on locating and evaluating grey literature as the main source for describing the current body of knowledge. Furthermore, the review scope was extended to incorporate pertinent studies on non-routine flight operations. Non-routine flying is a broader operational category, including positioning and ferry flights, instructional (training) flights, customer demonstrations, end-of-lease, and post-maintenance check flights (FAA, 2008a, 2016). This research study focuses on the safety of post-maintenance check flight operations.

The thesis adopted *maintenance check flight* (MCF) as an inclusive term per the European definition (see the list of terms in the previous chapter), covering all non-routine operations with a similar mission objective, such as post-maintenance test flights, functional check flights, post-modification evaluation flights, or acceptance flights. It is important to highlight that the term *check flight* is also used in the context of flight crew competency verification, however, those flight operations are not considered in this study.

EASA rules require a maintenance check flight, whenever prescribed in the aircraft maintenance manual (AMM) or other approved data; in the operator’s continuing airworthiness management system; or requested by the maintenance organisation for fault isolation, troubleshooting, or verification of successful defect rectification (Regulation (EU) 2019/1384, 2019). FAA guidance differentiates between two levels of complexity: functional and operational check flights. In FAA terminology, a functional check flight refers to an in-flight functional evaluation of the aircraft and its systems to a test standard, whereas an operational check flight reflects an in-flight verification of prior maintenance action or ongoing troubleshooting steps. The study incorporates learnings from both functional and operational check flight investigations (FAA, 2002).

Whilst the airline industry largely recognises that non-routine flying carries an elevated risk level, the actual magnitude associated with maintenance check flights is not that well understood by all stakeholders (FAA, 2016; FSF, 2011a; Poprawa, 2015). Available literature suggests a significant number of threats unique to non-routine flight operations which will be explored in this chapter.

The high-profile check flight and certification flight examples mentioned in the previous chapter put the spotlight on an existing safety problem that involves all forms of non-routine flying. Major check flight investigation reports delivered findings that pointed to similar systemic failures induced by airworthiness authorities, airframe manufacturers, airlines, and maintenance organisations. The realization that underlying systemic issues are not isolated to a particular airline, airplane model, or airworthiness system, led to an industry-wide consultation and change process initiated at the 2011 flight safety forum in Vancouver (FSF, 2011b). Thus, this chapter focuses on key safety events and investigations before and after the 2011 airline safety forum, the safety response adopted by the industry, and regulatory rules and accident investigation standards that framed the safety response. In addition, the review briefly discusses the changing work environment and new types of potential breakdowns introduced by modern airliners.

Specifically, the chapter is organised into six sections: a) the research methodology applied to the scoping review, b) the safety response adopted by the industry, c) a comparison between line and check flight operations, d) the evolution of highly automated airliners, e) relevant aspects of human decision making, and f) accident investigation standards. The chapter concludes by revisiting the key literature gaps identified by the review.

2.2 The Review Methodology

2.2.1 An Evidence-based Approach

As outlined in the previous section, there are very few research articles on the topic of maintenance test flight safety. Snyder (2019) highlights that for newly emerging topics, the primary purpose of the literature review is to create preliminary conceptualizations, rather than review old models. This type of literature review requires a more creative collection of data, not a synthesis of previous research.

Cooper et al. (Eds.) (2009) recommend a search strategy as a key element for finding relevant research evidence, including multiple iterations and phases needed to understand the topic and associated fields. They also highlight the increasing importance of searching grey literature when synthesizing scientific literature. Here they refer to the alternate definition proposed by McKimmie and Szurmak (2002, p. 72) that "grey literature consists of materials not identifiable through a traditional index or database, thus lacking standard bibliographic access". This broad definition includes government reports, safety bulletins, working papers, and conference proceedings that were instrumental for this study.

Landa et al. (2011) propose scoping reviews, a repeatable and evidence-based approach, for the first stage of a research project to ensure that further research is beneficial in the selected area. They suggest that an evidence-based approach, rather than one based on expertise alone, can better control the review and mitigate the introduction of bias at the start of the research project. The researcher's cognitive bias can distort the selection of material that can skew the subsequent enquiry and may even guide the study toward expected findings. The scoping review protocol includes a verification mechanism that repeats sub-selections and compares new abstracts with the original set.

This research study employed Levy and Ellis' (2006) concept-centric approach to the scoping review. The process follows a traditional data processing model (input-processing-output logic) common in engineering studies. Locating quality scholarly databases was aided by UniSQ Library experts. Meaningful keyword selection was guided by preliminary constructs, concepts, and theories uncovered during the initial search phase (Xiao and Watson, 2019). As articles were progressively reviewed, the depth of the review was further increased by exploring additional theoretical constructs and models, such as systems, organisational, information processing, and socio-technical theories. Extending the review relied on backward and forward searches for authors and topics, a technique introduced by Webster and Watson (2002) (Levy & Ellis, 2006, p. 189).

The structure and presentation of the literature review followed the general guidance and examples provided by Lunenburg and Irby (2008). The next section provides more details about specific keywords, search results, and the rationale behind embedding a systematic review of government archives for investigation records in a later chapter of the study.

2.2.2 Keyword Search and Evaluation

The author started the search by using the keywords "maintenance test flight", "functional check flight", "non-routine flight", "non-revenue flight" and "post-maintenance check flight" in literature resources recommended by the UniSQ Springfield Campus Library team. The search engines and databases included Google Scholar, UniSQ Library Indexed sources (Web of Science, EBSCOhost, Engineering Village, etc.), IEEE Xplore Digital, CiteSeerX, DOAJ, WorldCat, OpenGrey, and ProQuest. For each result, the title was evaluated to determine preliminary relevance. If the item appeared to discuss a topic relevant to check flying, or related to broader non-routine flight operations concepts, a full reference was obtained.

The initial search attempt delivered 813 results, including some duplicates. After the preliminary assessment, Google Search identified 24 relevant items, including 1 peer-reviewed article. UniSQ Library sources identified 11 relevant items, including 9 magazine articles, 1 FAA report, and 1 thesis. WorldCat identified 5 relevant items, including a book, 2 book chapters, a safety conference, and a magazine article. The book was later excluded from the review, as the full text could not be obtained. The author identified an additional 26 relevant items, including 1 peer-reviewed article, through backward and forward searching by author and references. The backward and forward searches were also effective in identifying 41 additional documents through following safety and regulatory initiatives, including loss of control, upset recovery, technical pilot training, flight test safety, and rule-making proposals. Overall, 78 source documents were included in the preliminary scoping review, before conducting a systematic review of government records (grey literature) for non-routine flight accident investigation reports (refer to Chapter 3).

Once the knowledge mapping study was completed (refer to Chapter 4 results), the search process was repeated by creating additional keywords from findings and safety recommendations embedded in maintenance check flight investigation reports. The concept-centric process led to selecting and reviewing additional topics related to the safety response, such as causal analysis, highly automated airliners, human error, and human decision making. Backward and forward search by authors and citations resulted in a research library that exceeded 500 sources by the time saturation was reached.

2.3 The Safety Response

High-profile accidents and incidents (e.g. 1996 DC-8 Narrows, 2008 A320 Perpignan, 2009 B737 Norwich, 2018 E190 Alverca) have drastically changed the airline industry’s perception of safety risk levels associated with maintenance check flying (AAIB, 2010b; BEA, 2010; GPIAAF, 2020; NTSB, 1997). In response, the FSF (2011b) organised an airline symposium in Vancouver to discuss check flight challenges with airline operators. The forum’s steering team included Airbus, Boeing, Bombardier, and Embraer. Before the symposium, most steering team members had recent exposure to serious *check flight* or *flight test* accidents, including several fatalities. The 2011 Vancouver forum was informed by major investigations conducted by the NTSB (United States), the BEA (France), and the AAIB (United Kingdom). Their findings, conclusions, and safety recommendations were instrumental in shaping the safety response later adopted and partially implemented by the airline industry (Airbus, 2015; FSF, 2011a).

This section reviews the regulatory changes and guidance introduced by the FAA and EASA, respectively, and safety action taken by manufacturers and airline operators.

2.3.1 FAA Framework

The FAA is an agency of the United States Department of Transportation, responsible for the regulation and oversight of civil aviation within the US, as well as the operation and development of the National Airspace System. Its primary mission is "The regulation of air commerce to promote its development and safety…", a somewhat conflicting mandate under the Federal Aviation Act (Federal Aviation Act, 1958, § 103(a)).

Part 91 Rules. Federal Aviation Regulation (FAR) Part 91 Subpart F 91.501 (b) allows operations that may be conducted under the rules in this subpart instead of those in parts 121, 129, 135, and 137 when common carriage is not involved. Operations allowed under 91.501 include ferry or training flights; aerial work; and demonstration flights (FAR Part 91, 2023).

Under Part 91.407 b) rules the aircraft does not have to be flown if, prior to flight, ground tests, inspection, or both confirm that the aircraft handling and its operation did not "appreciably change" during a maintenance event. Section 91.407(b) permits an air carrier two options to verify the accomplishment of the previously performed maintenance:

> (a) An operational check flight can be performed if any of the maintenance, preventive maintenance, rebuilding, or alterations

> may have appreciably changed the flight characteristics or substantially affected the aircraft flight operations…
>
> (b) The operator or air carrier may opt not to conduct an operational check flight by verifying the performance of the aircraft maintenance through ground checks or inspections, if appropriate (FAR Part 91, 2023, § 407(b)).

For this review, it is important to highlight that an airline can only operate a US-registered large and turbine-powered multiengine airplane under FAR Part 91 rules as a non-revenue flight, i.e. no common carriage is involved. When the rules require a check flight, only essential crew members can be carried on board until an appropriately rated pilot with at least a private pilot certificate performs the in-flight operational check of prior maintenance action.

As an outcome of the DC-8 investigation, the NTSB (1997) urged the FAA to introduce operating limitations and new training requirements for non-routine flights in FAR Part 121 Operating Requirements: Domestic, Flag, and Supplemental Operations. Instead of amending FAR Part 121, the FAA elected to change its guidance for airworthiness inspectors in a relevant Flight Standards Information Bulletin for Airworthiness instead, highlighting that most air carrier non-routine flights are conducted under Part 91 rules and not Part 121 (NTSB, 2003).

FAA NRFO Guidance (2002). The FAA (2002) bulletin provided the following national policy and guidance to airworthiness inspectors, aiming to assist airworthiness inspectors in ensuring that airline operators check flight procedures meet the requirements of the FAR:

> The term "functional check flight" indicates that the flight is conducted to a test standard. The term "operational check flight", as referenced in section 91.407(b), is a more appropriate term when referring to verification that the maintenance performed on the aircraft was accomplished to approved standards of repair and that the aircraft is operational (FAA, 2002, p. 1).

FAA NRFO Guidance (2008). In May 2008, the FAA (2008a) issued further regulatory guidance. It is noteworthy that it took the FAA over a decade to partially address the NTSB's (2003) concerns via an updated advisory material, leaving Part 91 rules unchanged.

The FAA (2008a) guidance introduced two groups for non-routine operations:

(a) Group 1 NRFOs involve the operation of the aircraft in accordance with the normal procedures of the flight manual; and

(b) Group 2 NRFOs that involve emergency, abnormal, or alternate procedures. These flights include any flight in which any system on the aircraft is de-powered to confirm that a back-up system is functional.

In line with the new FAA (2008a) position, an *operational check flight* falls into Group 1, and a *functional check flight* is a Group 2 operation. The FAA (2008a) also highlighted numerous Part 91 sections as applicable to air carrier NRFO flights, including pilot-in-command (PIC) responsibilities (as a final authority for airworthiness before, during, and after a non-routine flight), and the air carrier's responsibility for prescribing policies and procedures for authorizing and conducting non-routine flights, and flight crew qualification and training requirements in the air carrier manual.

FAA Safety Alert (2008). In December 2008, the FAA (2008b) issued a Safety Alert For Operators (SAFO), when responding to NTSB (2007) findings about the 2004 CRJ-200 fatal accident near Jefferson City. The SAFO clearly stated that the FAA (2008b) considered pilot error as a common factor contributing to non-revenue flight accidents. The SAFO cited, for example, the pilots' failure to adhere to SOPs and to operate the airplane within its performance limitations. The FAA (2008b) also urged Part 121 US operators to use data downloaded from Flight Data Recorders to monitor potential operational non-compliance and take corrective action against offending pilots.

FAA NRFO Guidance (2016). The Commercial Aviation Safety Team (CAST) is a joint FAA and industry safety working group formed in 1997. Having studied a series of accidents and incidents involving non-revenue flights, the CAST found that loss of the energy state or attitude awareness, and inadequately prepared flightcrew were some of the causal factors of non-routine occurrences. The CAST team recommended the FAA provide additional guidance to operators for conducting non-revenue flights.

In response, the FAA (2016) issued new guidance that highlighted the lack of SOPs and lesser degree of crew discipline driving an increased risk level. The FAA recommended that airlines introduce new risk controls through their operations manual, including specific mission risk assessments and briefings, and operating procedures closely following routine SOPs. The updated guidance also added task-specific training and supplementary checklist(s), should a flight program require operating with intentionally degraded systems.

Flight Simulator Fidelity. The NTSB's (1997, p. 40) final report about the Airborne Express crash identified the DC-8 simulator's low fidelity as a contributing factor, annotating that the simulator's benign flight characteristics provided the pilots with a misleading expectation of how the airplane responded in a fully developed stall. As a result, the NTSB (1997, p. 52) adopted a safety recommendation, urging the FAA to require improvements in reproducing stall characteristics to the maximum practical extent. In 1999, an internal FAA review concluded that airline pilots do not require training in advanced recovery manoeuvres, like deep stalls or stall breaks. Simulators did not need to provide fidelity into a stall beyond initial buffet or stick shaker activation. It follows that flight simulator manufacturers were not required to obtain accurate data about airplane stall characteristics from airplane manufacturers, even if aerodynamics performance data was available. The FAA response rejected the NTSB recommendation as not practical. The NTSB (2001) filed the original safety recommendation, citing an unacceptable response.

AOA Visual Indication. After the 1995 B757 accident near Cali, the NTSB (1996) adopted a recommendation that all transport-category airplanes present pilots with angle-of-attack (AOA) information in a visual format and that all airlines train their pilots to use that information to obtain the maximum possible airplane climb performance. The NTSB's (1997) final report on the 1996 DC-8 accident reiterated the same recommendation. In 2000, the FAA rejected the safety recommendation, arguing that AOA indicator installation is not warranted, as mandated Terrain Awareness and Warning Systems (TAWS) rendered the NTSB (1996, 1997) recommendation moot. In its response to the Safety Board, the FAA also noted that:

> The FAA's primary focus is to enhance the flightcrew's situational awareness well before encountering an emergency condition, so that maximum airplane performance need not be relied on during ground proximity escape maneuvers (NTSB, 2001, p. 2).

The FAA's argument is questionable. TAWS systems proved to be very efficient in reducing Controlled Flight Into or Toward Terrain (CFIT) accidents, but a TAWS system does not address the Safety Board's intent to maximize airplane climb performance under the circumstances encountered by the ABX DC-8 crew in 1996 (NTSB, 1997). Furthermore, the FAA's comment about situational awareness is confusing, as it appears to deny the dynamics involved in an in-flight emergency scenario. It also implies that pilot error lies behind the need for any escape manoeuvre which would require maximum airplane performance. The NTSB (2001) closed the safety recommendation, citing an unacceptable FAA response.

In 2013, the French Air and Space Academy provided an excellent summary of the unresolved issue surrounding the need for AOA indication in commercial airliner flight decks, when it observed that:

> Information on angle of attack is paramount to conduct of the flight, although the reliability of airspeed indications consigned it to the background for a while, until its comeback in the shape of the velocity vector. On commercial airliners it is absent for the

> moment probably because manufacturers were reluctant to add yet another instrument to the instrument panel and above all because airlines saw no need for it in spite of demands from professional organizations, the NTSB and the Flight Safety Foundation (Pinet & Buck, 2013, p. 68).

2.3.2 EASA Framework

Within the European Community, the European Union Aviation Safety Agency (EASA or *the Agency*) is an independent body concerning technical matters and has legal, administrative, and financial autonomy. EASA exercises implementing powers conferred by Regulation (EC) No 216/2008. Its primary responsibility is the effective functioning of the European civil aviation safety scheme to detect unsafe conditions and to take remedial measures, as appropriate (Regulation (EC) No 216/2008, 2008).

Rulemaking Task. In 2012, EASA (2012) proposed amendments to existing operational rules, having accepted earlier findings and recommendations from the BEA (2010) and the AAIB (2010b) for implementing consistent safety measures for non-commercial flights in the European regulatory context.

During the rulemaking process, EASA (2017) also recognised that airline operators used inconsistent terminology and procedures for non-commercial flights, there were no clear definitions and requirements in

EU-OPS for the different non-commercial operations, minimum criteria for the selection of flight crews to perform these flights were absent, and crews had to frequently improvise during critical flight phases.

New MCF Regulations. After a decade-long regulatory development process, the new EASA framework introduced two levels for post-maintenance check flight operations. Level A applies when the operator expects to use abnormal or emergency procedures in the airplane flight manual, or when testing of the correct functioning of a backup system or other safety devices is required. Level B applies to less complex flights, when Level A requirements do not apply. Under the new EASA rules, only essential crew members can be carried on board to conduct a Level A post-maintenance in-flight check (Regulation (EU) 2019/1384, 2019).

In support of Level A check flights in a complex motor-powered aircraft, the new regulations also require European operators to develop and document a test schedule, to have dedicated procedures or a check flight manual endorsed before conducting these operations, and to develop flight crew competency through simulator or flight training following a detailed syllabus. Furthermore, airlines must select pilots who meet minimum training, experience, and recency requirements, similar to a flight test environment. They also need to assign a task specialist, or an additional pilot, when the anticipated crew workload demands additional assistance; and have a risk management plan and appropriate risk mitigations agreed upon before conducting functional checks of any system or equipment. Importantly, the operator must treat the system or equipment under test as unreliable, planning adequate margins for a worst-case scenario (Regulation (EU) 2019/1384, 2019).

Flight Tests and Flight Checks. EASA non-commercial operational rules do not address certification or production test flights (Regulation (EU) 2019/1384, 2019). In the wake of the Perpignan accident, EASA (2008) released a Notice of Proposed Amendment (NPA) which distinguished flight tests and flight checks. The NPA stated that "during such flights there is a certain amount of unpredictability which does not happen in the case of check flights and acceptance flights." (EASA, 2008, p. 9).

Supported by evidence from multiple check flight investigations, the AAIB (2010b) highlighted that the same level of risk can exist during a flight test and a check flight, should unidentified defects be present or inappropriate crew action introduces additional risk, essentially rejecting the EASA (2008) position. The AAIB (2010b) urged EASA to achieve comparable minimum safety standards, through minimum crew proficiency and recommended crew composition and training requirements.

In 2018, the European Commission harmonized flight test crew qualification requirements and recommended that EASA Part-21 approved organisations involved in flight testing develop a Flight Test Operation Manual which is proportionate to the complexity of the airplane and the design and production organisational structure (EASA, 2018). In terms of check and acceptance flying, the new implementing rules adopted by EASA for non-commercial and specialised operations introduced minimum flight crew training and proficiency standards that are comparable to those recommended for lower risk flight test operations (Regulation (EU) 2019/1384, 2019).

2.3.3 Industry Response

Investigation reports released by the NTSB (1997), the BEA (2010), and the AAIB (2009, 2010a, 2010b, 2013) all highlighted the essential role airframe manufacturers played in preventing check flight safety issues. Their findings and recommendations primarily focused on ensuring that in-service non-routine operations were supported by appropriate instructions for continued airworthiness, including engineering and operations manuals, flight test schedules, and flight operations support on-demand.

Multiple models of the Airbus brand were exposed to increased public scrutiny due to past non-routine accidents and serious incidents, and as a result, Airbus' (2015) approach to proactively exceeding current regulatory requirements was in sharp contrast with the slow and controversial response from aviation regulators. A dedicated technical pilot training program, the In-Service Aircraft Technical Flight Manual released to airlines on request (with liability waivers), an updated AMM subject for every model to reflect a revised Airbus non-revenue flight policy, and regular safety communications are examples of the improved product support measures introduced by Airbus.

Boeing was somewhat more subdued in its response to the check flight safety problem. Boeing remained firmly focused on reassuring its existing and potential customer base regarding the safety of its models (FSF, 2011a). To put the Boeing response into context, it should be noted that from their product line, only the 737 family was exposed to intense scrutiny when the industry turned its attention to high-profile check flight incidents. Whilst Boeing complied with all FAA and EASA adopted safety recommendations – through updated AMM instructions and generic Flight Test Schedules released for each model – Boeing’s approach can be summed up as reactive, awaiting the outcome of the FAA and EASA regulatory development process outlined in the previous sections.

As mentioned earlier, the Functional Check Flight steering team (comprising Airbus, Boeing, Bombardier, and Embraer), organised the 2011 airline symposium to discuss check flight challenges with the operators. Manufacturer presentations at the Vancouver symposium echoed the *be prepared* safety message originally introduced by Airbus in 2008, shortly after the Perpignan tragedy (FSF, 2011b; Nelson, 2008a).

Other airframe manufacturers also highlighted their *planning and preparation* for mitigating the risk of experimental test flights and customer acceptance flights, however, their key message remained firmly focused on reassuring the symposium that check flights pose no additional safety risk, if only prepared and managed properly by airline operations. Having accepted BEA (2010) and AAIB (2010b) safety recommendations, generic check flight schedules became available for in-service use, with the provision that airlines remained fully responsible for adapting test schedules to local airworthiness requirements, and accounting for any configuration differences within their fleet (FSF, 2011a).

After the symposium, the FSF Foundation and the Functional Check Flight steering team continued their safety awareness campaign, which resulted in the publishing of a Functional Check Flight Compendium, originally intended as a guidance document for operators (FSF, 2011a). Despite being generic, the compendium contained very practical recommendations, albeit echoing the plan and prepare philosophy. During the regulatory development phase, EASA (2017) proposed to accept the compendium as an alternative means by which the airline could safely prepare for conducting a check flight. Unlike the FAA (2016), the final

EASA position adopted in 2019 does not endorse the 2011 compendium as an acceptable means of compliance (Regulation (EU) 2019/1384, 2019).

2.3.4 Section Summary

Within a decade after the NTSB (1997) published its final conclusions about the 1996 DC-8 Narrows investigation, a series of additional tragic accidents and serious incidents served as a reminder for the airline industry that check flight safety improvements were overdue. In line with ICAO Annex 13 obligations, national transport safety agencies conducted their investigations and adopted a range of safety recommendations, urging relevant airworthiness authorities to *fix* the uncovered airline safety deficiencies. To borrow a term from system safety, the response can be summed up as a *fly-fix-fly approach* to safety (Leveson, 2001; Roland & Moriarty, 1990).

Based on a chronological review of regulatory changes and safety action taken by key stakeholders, the review found that:

- Most safety recommendations were accepted by the responsible airworthiness authorities; however, extensive implementation delays were incurred when creating new operational safety standards.

- The extensive regulatory delays are symptomatic of a low safety priority afforded to the problem, which poses the question of whether the airline industry fully understands the actual level of operational risks involved in check flying.
- The manufacturers’ attitude conflict, when describing check flight risk exposure as not comparable to that of experimental flight test operations, was not challenged by the FAA or EASA.
- Airframe manufacturers were active in communicating the safety risk to their airline customers, but their *plan and prevent* awareness campaign was firmly focusing on air carriers’ in-service operations and responsibilities.

The FAA and EASA appear to have accepted the manufacturers’ proposition that check flight safety is an in-service problem and should be treated as such (FAA, 2016; Regulation (EU) 2019/1384, 2019). The problem with treating the safety issue solely as a continuing airworthiness problem is that any potential linkage to underlying design or certification issues is discarded.

The following section reviews the literature for independent research articles about check flying and non-routine flight operations, more broadly.

2.4 Line and Check Flight Operations

2.4.1 Knowledgeable Technical Pilots

Veillette (2009) conducted an extensive review of post-maintenance test flight accident reports and NASA Aviation Safety Reporting System records and revealed a significant number of threats unique to non-routine operations. His review identified the extra cockpit workload (because of the abnormal procedures), abnormal crew coordination, and a likely aircraft abnormality (or an inoperative component), as common threat themes across occurrence reports. He concluded that coordination and communication between maintenance engineers and knowledgeable NRFO pilots was key to safe check flight operations.

When Veillette (2009) coined the term *knowledgeable NRFO pilot*, he envisaged a specially trained and skilled airline technical pilot, one who received additional training, who can coordinate and communicate well (both on the ground and in the cockpit), someone who is thoroughly knowledgeable about aircraft systems, and alert to warning signs about deteriorating aircraft behaviour. Furthermore, in his view, the ideal technical pilot knows how their aircraft is likely to react and is fully competent to safely recover the aircraft to stable flight. With reference to historical complex loss of control scenarios during fatal maintenance test flight accidents, Poprawa (2015) arrived at similar conclusions about the specialist piloting skills required for airline technical pilots. However, his study also raised concerns about whether existing airline pilot training standards were adequate for performing check flight operations in a safe manner.

Airbus (2015) recommends similar professional and personal traits when recruiting pilots and engineers for the *checking community*. According to their guidance, the following four pillars are essential for any technical pilot candidate:

1. Detailed knowledge of the aircraft, the theory, and the role, combined with an inquisitive mind. Airbus (2015) would expect technical pilots to keep asking questions until they receive an answer that is both *right* and *makes sense*. This latter part of the recommendation is problematic, though. It is impossible to qualify an answer as *right* without knowing the outcome, a common trait of hindsight bias (Fischhoff, 1975).

2. The ability to observe, interpret, analyse, and properly communicate.

3. In terms of aptitude, Airbus (2015) found that both check pilots and engineers should excel in Crew Resource Management (CRM) and the ability to listen. The guidance also underlines that the candidate’s personal integrity should be valued above everything else, more of a moral quality than natural ability.

4. The fourth pillar is experience. Airbus (2015) urged airline operators to look beyond logbook hours, though, alluding to the fact that pilots with thousands of hours on the line may be totally unsuited to conduct check flight missions.

The FAA did not mandate technical pilot training and qualification standards for the purposes of non-revenue flight operations. FAA (2008a) guidance issued in response to the 1996 DC-8 crash recognised the need for additional training for Group 2 NRFO (functional check flight) purposes, even if the FAA (2008a) maintained that specifically trained pilots were "qualified to simulate emergencies and expect the unexpected" (p. 4). *Unexpected scenarios* cannot be trained for, and as such, it is difficult to interpret this part of the FAA guidance.

In the FAA framework, Group 1 NRFO flights (operational check flights) may be conducted by a Part 121 carrier's line pilot, without any additional training. According to the FAA (2008a), every Part 121 PIC is trained and qualified to operate the airplane in all normal procedures and to recover the aircraft should there be any problems during the flight. FAA (2008a, 2016) guidance clarifies that Group 2 NRFO missions (functional check flights) can benefit from providing task-specific training and briefing to crew members, especially when in-flight checks call for operating the airplane with intentionally degraded systems. In summary, the FAA (2008a, 2016) position indicates a view that check flights are not comparable to test flights and the crews conducting non-routine operations need not be qualified test pilots, owing to very different risk profiles. Furthermore, the FAA (2008a, 2016) guidance echoes the *plan and prepare* approach, stating that by following their methodology, crews can plan for unexpected occurrences during a non-routine mission and should they encounter in-flight problems, previously endorsed procedures should enable a safe landing.

Contrary to the FAA position, EASA decided to mandate minimum flight crew competency standards (Regulation (EU) 2019/1384, 2019). EASA crew requirements for a Level A check flight (broadly equivalent to a Group 2 NRFO in FAA terms) in a commercial jet can be briefly summed up as:

- minimum flight experience is 1000 hours in the same category, of which a minimum of 400 hours as a PIC in a complex motor-powered aircraft, and at least 50 hours on type.
- the candidate has completed a technical pilot training course, including ground and simulator elements. The course is valid for all types; however, first check flight must be conducted as a copilot or observer if the training was provided in a simulator only.

- recency requirement of a minimum of one check flight in the preceding 36 months, one check flight as a copilot/observer, or one simulated check flight as a PIC, if out of currency, and
- pilots holding a flight test rating are eligible for full credit for the technical pilot training course, provided they obtained initial and recurrent CRM training which meets the intent of the regulation.

EASA also adopted industry best practices when mandating that airline operators should assign additional technical specialist(s) or an additional pilot as a crew member, considering the anticipated workload based on the prescribed test program (Regulation (EU) 2019/1384, 2019). Similar mixed crew complements have been in place at a major European airline group for many years, which were likely to have contributed to an excellent operational check flight safety record.

To mitigate safety risks associated with non-routine operations, the international airline pilot community adopted a similar position regarding specific training standards. For example, a more recent position paper released by IFALPA (2019) highlights that pilots need to obtain the necessary experience through simulator sessions, observation flights, and/or crew pairing with pilots familiar with non-routine operations.

Poprawa (2015) identifies the lack of regulatory standards for technical pilot qualification and flight simulator fidelity as key challenges in the context of check flight safety. Flight simulator fidelity (especially loss of control, upset recovery, proper motion cues, and the difficulties involved in developing representative simulator data packages by non-linear aerodynamic modelling) is a well-explored topic in the literature (International Air Transport Association [IATA], 2016; Lande, 2016; Simulation of Upset Recovery in Aviation, 2013). It is beyond the scope of this review to explore complex simulator training in further detail.

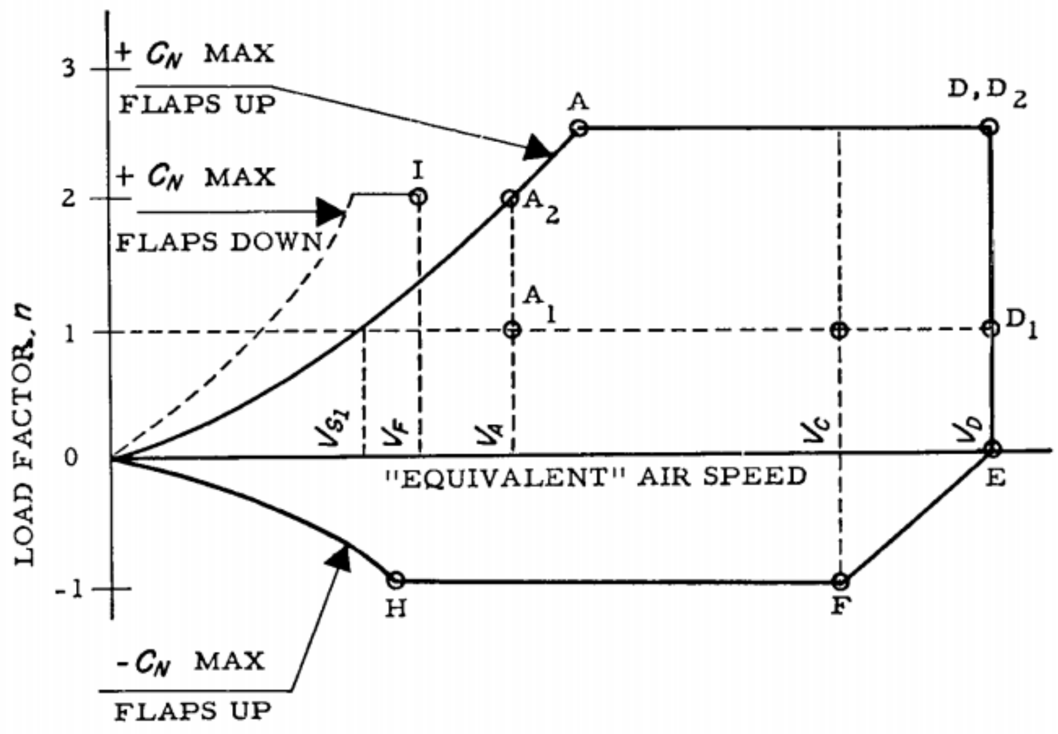
In Veillette's (2009) opinion, extra simulator training has some safety benefits, but not necessarily so in the context of check flight operations. While he recognises that simulation had markedly improved since the DC-8 accident, he maintains that even modern simulators were not intended to replicate flight operations at the edge of the aircraft's performance envelope. He cites the 1996 Narrows accident as one example in which the control techniques learned in the simulator aggravated the in-flight upset experienced by the pilots.

2.4.2 Line and Check Flying

Manoeuvring Envelope. FAR Part 25 rules provide the primary specification for any new and modified transport category aircraft design, and the main reference points for any design engineer involved in aircraft development in an FAA context. Embedded within Part 25 rules is a fundamental structural specification which must be fully complied with by any transport airplane design (FAR Part 25, 2023).

Figure 1

V-n Diagram for Transport Airplanes



Note. The V-n diagram was reproduced from FAR Part 25.333 (FAR Part 25, 2023).

The manoeuvring envelope in Part 25.333 defines the fundamental in-flight characteristics of a transport aircraft, which must be met at each combination of airspeed and load factor represented by the envelope, including the area enclosed, as well as the boundaries of the envelope. The manoeuvring envelope is commonly referred to as the *V-n diagram*, with reference to the required airspeed and load factor combinations. The V-n diagram also serves the purpose of showing compliance that the design meets or exceeds structural strength requirements imposed by Part 25 (see Figure 1).

Airline pilots spend most of their time during line flying around the intersection of V_c *(the design cruise speed)* and *load factor* $n_z = +1.0$, which translates to a normal positive +1.0 g flight, provided wind gusts or other minor accelerations during the cruise segment are not accounted for. When landing, the crew slows the airplane from V_c, which means that the airplane is moving to the left on the envelope and will remain just to the right of the curved dashed line (depicting minimum speed with landing flaps extended) for the last few seconds of the flight, before touching down. During a take-off sequence, the airplane moves in the opposite direction on the envelope, close to the line representing $n_z = +1.0$. When the airplane lifts off from the ground, the relevant point is a bit above that same dashed line. In normal turns, up to 30 degrees Angle-of-Bank (AOB), the load factor would similarly be slightly above the line representing $n_z = +1.0$. Other than unique simulator sessions, the rest of the manoeuvring envelope remains unexplored, and potentially never experienced, by an airline pilot (IATA, 2016; Pinet & Buck, 2013).

Contrary to line pilots, experimental test pilots are involved in exploring the edges of the design envelope during certification flights. In this context, the term *experimental* refers to the inherently higher risk involved in testing a new design or an airplane that incorporated major changes to the type design, prior to a type certificate or supplemental type certificate being awarded (FAR Part 91, 2023). To paraphrase an experimental test pilot, who flew aerial displays of transport category aircraft, the most remarkable part of the job was the fact that he could go from one extreme end of the flight envelope to the other and back again in only 30 seconds. In other words, the airplane might be approaching stall, only to exceed V_{MO} (maximum operating speed) 15 seconds later (IATA, 2016, p. 8).

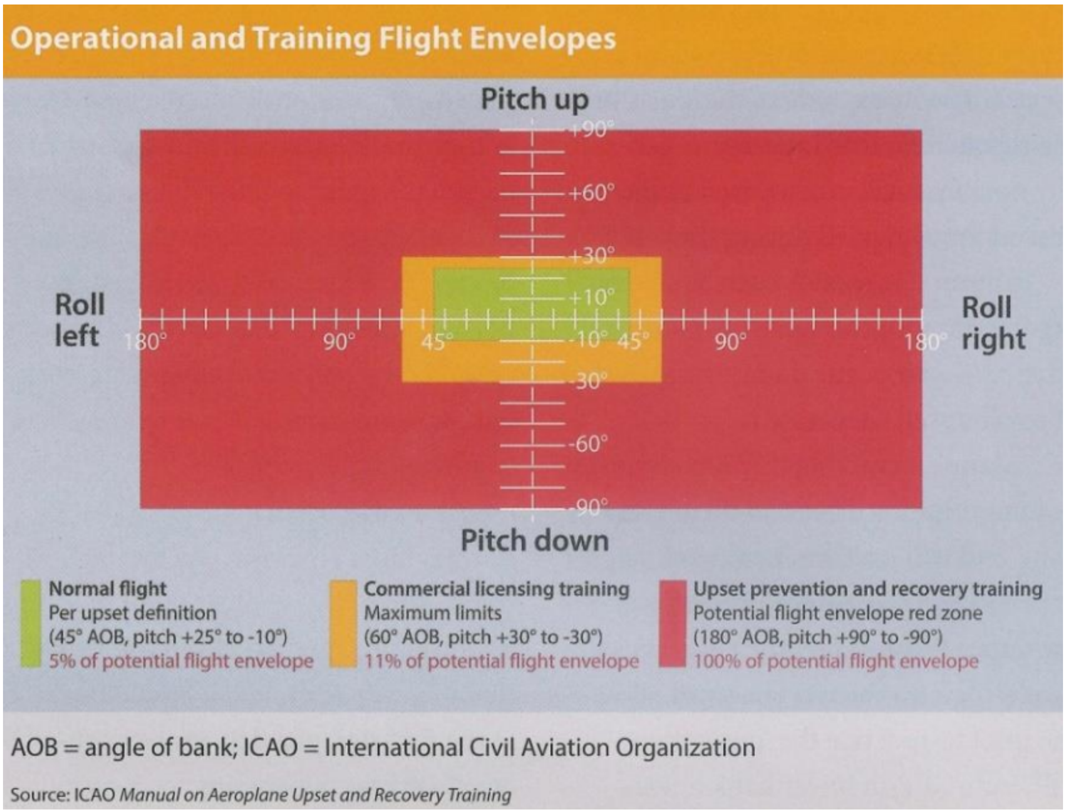
During an uneventful check flight, the pilot will remain within, or very close to, the region of normal line flying. Should a situation develop during the test sequence that is not normally experienced by airline pilots, the airplane may end up at any other point on the V-n diagram and the pilot will need to recover from that unexpected state. The responsibility of an airline technical pilot falls in a grey area, somewhere between certification and normal line flying.

Operational and Training Envelope. The ICAO (2014) Manual on Aeroplane Upset and Recovery Training offers another way of looking at the differences between line flying and the role of a technical pilot during a check flight. Type rating programs focus on a narrow band of the operational envelope experienced by line pilots (ICAO, 2014, p. 65). The difference between operational and training flight envelopes perfectly illustrates the difficulties in coping with an in-flight upset scenario (see Figure 2).

Lande (2016, p. 5) succinctly summed up the challenge when he posed the question "How can we expect pilots to cope with a flight upset situation involving more than 90° of bank, when they have only seen 60° during their training and are normally limited to 30° during normal operations?". Lande's observation is directly relevant to any airline pilot who is tasked to command a check flight without any additional training. The pilot may have a wealth of experience in a line environment, which is normally limited to 30 degrees AOB and relatively benign pitch attitudes, and probably experienced scenarios up to 45 or 60 degrees AOB and +/- 30 degrees pitch during simulator training, illustrated by the innermost and middle rectangles in the envelope, respectively. When commanding a check flight, the airplane may easily end up anywhere within the outermost zone of the envelope, an area completely unknown for a line pilot, which in turn poses an unmitigated safety risk during non-routine operations (Lande, 2016; Poprawa, 2015).

Figure 2

Operational and Training Flight Envelopes



Source: ICAO *Manual on Aeroplane Upset and Recovery Training*

Note. The diagram was reproduced from Lande's (2016) conference presentation to the International Society of Air Safety Investigators.

2.4.3 Comparing Routine and Non-routine Safety

Several international organisations, aircraft manufacturers, and insurance companies publish reliable safety reports about routine commercial flight operations such as ICAO (2022), IATA (2020), Airbus (2022), Boeing (2022), or Allianz Global Corporate & Specialty (2014).

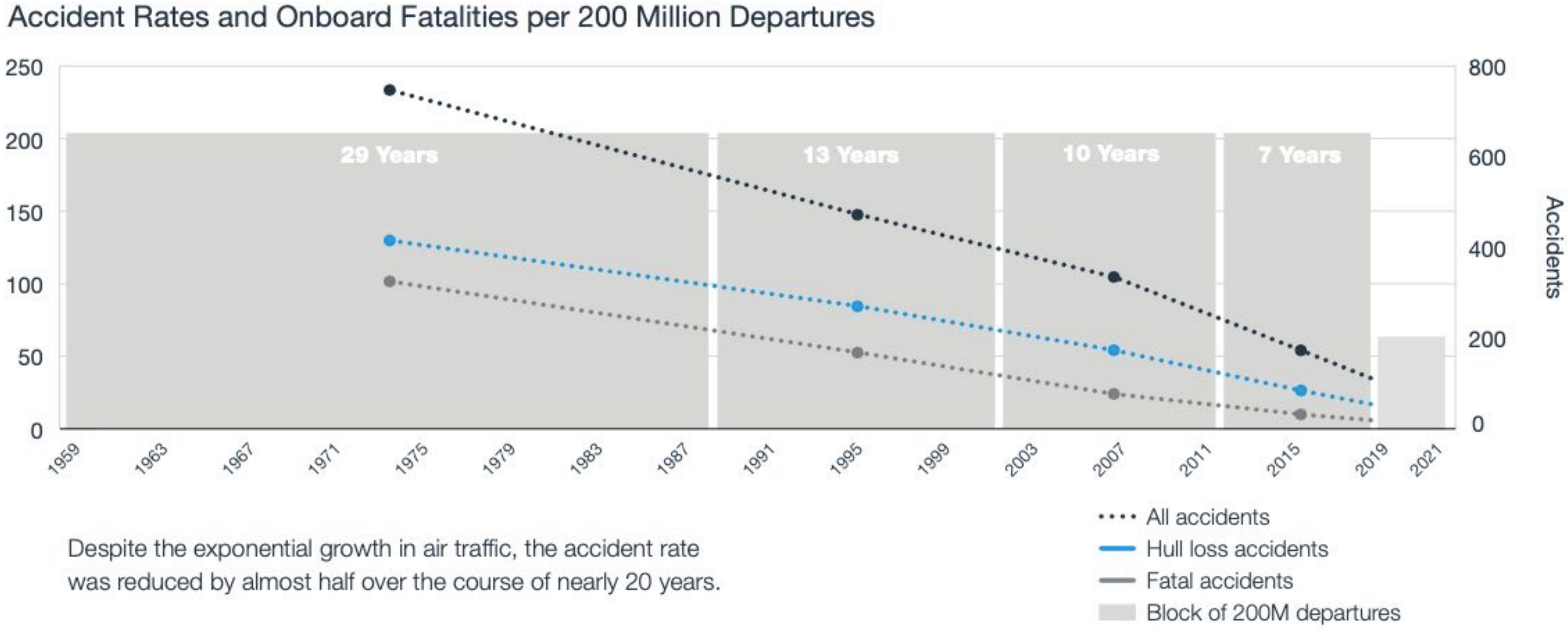
The airline industry experienced unabated exponential growth before the most recent global health emergency. Figure 3 depicts accident rates against blocks of 200 million departures, including hull loss and fatal accident rates. As noted by Boeing (2022), accident rates only twenty years ago were almost double the most recent statistics, a remarkable achievement by airframe manufacturers, operators, service providers, professional organisations, and airworthiness authorities involved. Amalberti (2001) refers to civil aviation as an example of today's ultra-safe macro-technical systems. Current safety levels in the nuclear industry or high-speed train networks may also serve as representative examples of similar ultra-safe systems.

In the context of routine operations, there was a clear and distinct improvement in safety outcomes between successive airliner generations, as illustrated by accident rates over the life of major technological improvements (see Figure 4). Third-generation commercial airliners equipped with electronic cockpit displays (glass cockpit) have been in service since the early 1980s and have also benefited from embedding digital technology, like Flight Management Systems (FMS) and Terrain Avoidance and Warning Systems (TAWS). The improved navigational and terrain avoidance capabilities were essential in sharply reducing CFIT rates (Airbus, 2022; Boeing, 2022). Fourth-generation FBW technology-enabled flight envelope protections have also demonstrated some remarkable safety improvements, offering additional safety margins and automated protections during routine airline operations. In terms of both fatal accidents and hull losses, the latest generation has the lowest accident rates, at a stable average of 0.1 to 0.2 accidents per million departures. This reflects a 50 percent reduction when compared to third-generation airliners (AGCS, 2014; Airbus, 2022; Boeing, 2022).

Most commercial aviation safety reports exclude occurrences during non-routine flying operations. As noted earlier, Boeing's annual statistical summary is an exception, as it provides a tally of hull losses and fatal accidents for maintenance test, ferry, positioning, training, and demonstration flights. According to the latest Boeing (2022) summary, between 1959 and 2021, the worldwide commercial fleet suffered 124 accidents during non-routine flights and 1,981 accidents during routine sectors. Over a ten-year period between 2002 and 2011, there were 13 non-routine and 391 routine accidents, including 3 and 76 fatal accidents, respectively (Boeing, 2012). The following decade registered improved safety statistics, where none of the 3 non-routine accidents were fatal and only 36 of the 304 routine accidents led to onboard or external fatalities (Boeing, 2022). Owing to difficulties in accounting for the actual number of non-revenue flight sectors, non-routine accident and hull loss rates are not readily available in published literature.

Figure 3

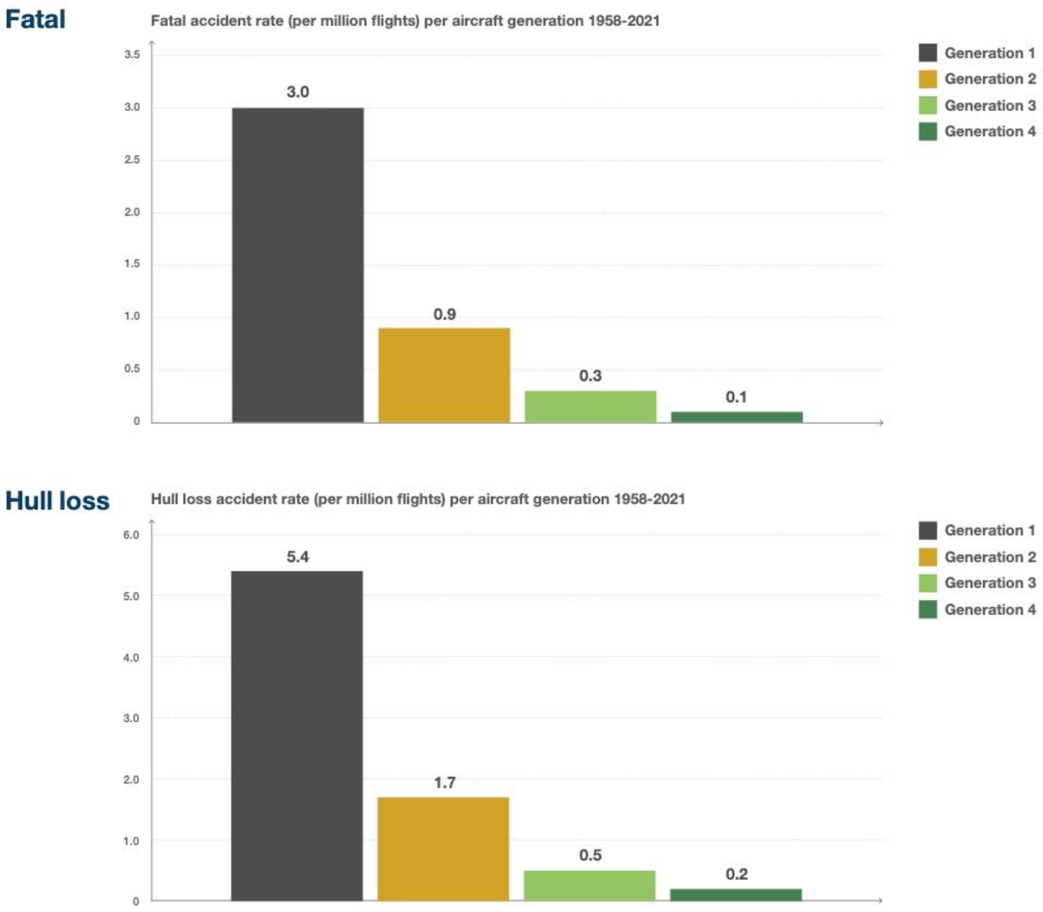
Accident Rates and Onboard Fatalities



Note. The source of the diagram is Boeing's (2022) Statistical Summary of Commercial Jet Airplane Accidents.

Figure 4

Fatal and Hull Loss Accident Rates per Airliner Generation



Note. The source of the diagram is Airbus' (2022) Statistical Analysis of Commercial Aviation Accidents 1958-2021.

Independently from this review, Poprawa (2015) found that maintenance test flights remain "one or two orders of magnitude more risky" than routine flying (p. 1). He estimated that the ratio of check flights and routine sectors is approximately one maintenance test flight for every 1,200 to 1,800 revenue sector, assuming a low aircraft utilization rate. Based on those conservative estimates, and the number of fatal accidents outlined in the previous paragraph, a single fatal check flight accident in a ten-year period translates to a risk level that is at least 33 to 50 times higher than routine line operations. Other than Poprawa's (2015) findings, the literature provides no guidance in terms of actual check flight safety risk levels. With reference to Amalberti (2001), a magnitude higher operational risk level would indicate that current industry practices had optimized check flight operations as a *regulated system*. Since incident volume monitoring (i.e. calculating accident rates) is less relevant for regulated systems, direct analysis of check flight incidents is more applicable when seeking additional safety improvements.

With very limited non-routine accident data available in the literature base, it is not feasible to compare check flight and routine safety outcomes without further research. Reasonable estimates indicate that global safety levels achieved during check flying should not be treated like ultra-safe commercial line operations. Therefore, it is not clear whether airlines, regulatory bodies, manufacturers, and safety investigation boards fully understand the actual level and nature of operational risks involved. If systemic operational hazards cannot be fully characterized and exposed, airline safety outcomes remain vulnerable to inadequate risk management, including unsuitable, missing, or partially effective risk controls. This is a critical gap in the literature.

2.4.4 Section Summary

This section reviewed the literature for independent research studies and journal articles. The few available studies corroborate findings in landmark check flight investigation reports. A key underlying theme is the role played by knowledgeable technical pilots in preventing (or contributing to) future accidents. Safety concerns are clearly linked to elements of ongoing aviation safety research and airline training development, including upset recovery, loss of control, and flight simulator fidelity studies. The regulatory framework does not fully align with available research evidence and industry best practices, including crew selection, training, and crew composition. Regulators are aligned with the plan and prepare approach introduced by airframe manufacturers, tacitly endorsing check flight safety as an operational airworthiness issue.

Owing to the lack of research articles, the literature review was broadened to explore differences between line and check flying and available safety statistics for routine and non-routine flight operations. Routine commercial flying maintains an ultra-safe operational record. In contrast, check flight operations carry a magnitude higher risk, currently optimized as a regulated system. It is not clear whether an adequately trained and authorized crew would be able to meet the novel and unexpected challenges posed by check flying highly automated airliners.

The scoping review confirmed that it was not feasible to compare historical routine airline safety statistics with the very limited data available for check flight accidents. This shortfall indicates a critical literature gap which warrants further research. In line with Levy and Ellis' (2006) evidence-based methodology, the literature review was widened, looking for novel system safety challenges created when expert operators interact with highly automated machines, including the role of human decision making, followed by evaluating the suitability of accident investigation models to account for those interactions.

2.5 Highly Automated Machines

Initial iterations of the scoping review confirmed that modern commercial airliners are not immune from suffering safety losses during check flight operations. To a large extent, post-maintenance test flights are concerned with performing an in-flight validation of automated systems, functions, and flight deck effects. The mission objective is different from routine line operations and raises questions about potential new threats and challenges. This section explores the challenges created by the evolution of flight systems, the changing flight deck interface, and advanced automation in a system safety context.

2.5.1 Flight Systems Development

From the early flying machines to today's sophisticated airliner generation, stability and control have always been crucial design considerations. During the "jet's awkward age" the need for redundant, irreversible power controls and electronic stability augmenters was evident but non-electronic stability augmentation was the only reliable option available for designers. Highly swept wings, high Mach numbers, transonic, high-altitude flying, and higher wing loadings led to rapid developments in jet airplane stability and control (Abzug & Larrabee, 2002; McRuer & Graham, 2004).

The first commercial jets entered service with early autopilot systems, replaced by more elaborate auto-flight and auto-throttle systems in the mid-1960s. The next major development milestones were electronic cockpit displays, FMS for improved navigation, and terrain avoidance and warning systems to address CFIT accident rates. While digital FBW control system technology goes back to the 1970s, complex design, simulation, ground and flight testing, and certification issues had to be addressed, before the first digital FBW airliner entered commercial service in 1988 (Abbott, 2000; Airbus, 2022; ICAO, 2013).

The rapid pace of flight systems evolution can be best illustrated by some of the more successful designs. The Boeing 747, Lockheed L-1011, and Douglas DC-10 utilized various new implementations to provide fully automated landing functions. The Boeing 747 used triple redundant analogue computers. The L-1011 used digital computers in a dual-dual architecture. The DC-10 used two identical channels, each consisting of dual-redundant fail-disconnect analogue computers for each axis. The first electrical control system for civil aircraft was installed on the Concorde. It was an analogue, full-authority system for all control surfaces (Hitt & Mulcare, 2000). The Airbus A320 family introduced a full-time digital FBW flight control system. The basic architecture paved the way for modern fault-tolerant FBW design to meet stringent dependability requirements in terms of safety and availability (Briere & al., 2000; Briere & Traverse, 1993; Favre, 1994). In a similar vein, the heart of the Boeing 777 FBW design is a triple modular redundancy concept integrated with an N-version dissimilarity approach. For example, the flight control computer architecture uses a 3 by 3 matrix of 9 processors of 3 different types, built with a fail-passive electronic arrangement (Yeh, 1996, 1998, 2014).

For the purposes of this research project, there was no need to consider detailed system architecture solutions. The thesis adopted Airbus’ (2022) description and ICAO's (2013) definitions when a reference is made to successive airliner generations, as follows:

Generation 1 – Early commercial jets from 1952. Analogue cockpit and early auto-flight systems (B707, BAC-111, Caravelle, Comet, Convair, DC-8, Trident, VC-10).

Generation 2 – More integrated auto-flight from 1964. Better autopilot and auto-throttle systems (A300, B727, B737-100/200, B747-100/200/300/SP, Concorde, DC-9, DC-10, F-28, L-1011).

Generation 3 – Glass cockpit and FMS from 1980. Electronic cockpit displays, improved navigation performance, and TAWS (A300-600, A310, B717, B737 Classic/NG/Max, B757/767, B747-400, Embraer ERJ models, F-70/100, MD-11, MD-80, MD-90).

Generation 4 – Digital FBW from 1988. FBW technology enabled flight envelope protection (A220 models, A320 family, A330/A340, A350, A380; B777, B787, Embraer E-Jets).

2.5.2 Flight Deck Evolution

Traditional planning for embedding new technologies in complex systems took an automation-centred approach (Abbott & Rogers, 1993; Billings, 1996). This approach was built on the premise that a flawless design process can produce an output that handles all scenarios it may encounter in an operational environment. Starting with the classic human factors issue of *allocation of function* between humans and the machine, automation can be defined as the machine agent executing a function previously carried out by a human operator. As automation increases, greater efficiency, increased flexibility, and better safety outcomes are expected as a result (Billings, 1996). However, this traditional approach contradicts the real-life experience and safety concerns raised about cockpit automation (Parasuraman & Mouloua, 1996; Parasuraman & Riley, 1997, Sarter & Woods, 1992, 1997; Wiener & Curry, 1980; Woods & Sarter, 2000).

Traditional flight deck design was evolutionary and technology-driven, which offered a relatively low-risk approach and a set of design requirements which are reasonably straightforward to implement (Funk, 1991). Wiener (1989) observes that limiting the design scope to incremental changes, such as correcting in-service issues by adding one box at a time, was a common trait of the conventional approach to cockpit automation. The designer had to focus on isolated tasks when observing crew behaviour, which was then developed into a design solution (Nakamura et al., 2013). The systems engineering approach introduced the concept of a multiple task environment (Damos, 1991). The new approach was supported by Funk's (1991) Cockpit Task Management theory, which was sufficiently intuitive and well understood by airline pilots, and an error taxonomy based on multiple NTSB accident reports (Sheridan, 1988). The underlying objective was to introduce countermeasures to human error through an improved cockpit design, treated by the designer as a pilot vehicle interface issue (Chou, 1991).

During the development of an advanced flight deck for a commercial airliner, the initial step is to define the primary flight deck functions needed to support the mission goals of the vehicle. Automation design must work cooperatively with human operators in pursuit of those mission objectives (Abbott & Rogers, 1993). For a routine mission, the overriding objective is to move passengers and cargo from airport gate to airport gate and to do so safely and efficiently. For a check flight, the goal is to verify the integrity and functionality of safety-critical systems in flight. These mission goals are not mutually exclusive, however, there is no evidence that primary cockpit functions can safely and efficiently support check flight operations.

Sarter and Woods (1992) observe that when automated flight controls (the autopilot) and automated flight management systems (FMS and FBW) were introduced, the distance between the pilot and direct control of the aircraft grew even further. Coupled with an increasing systemic complexity, this change led to the pilot understanding less about automation states and behaviours. The new technology also raised safety concerns, especially in unexpected or emergency conditions. To date, considerable research effort has been dedicated to understanding the implications of advanced automation, like the introduction and acceptance of FMS systems, however less attention was paid to cockpit warnings (e.g. stall warning, ground proximity warning, traffic collision systems, etc.) or new information systems (like ATC datalink, traffic information displays enabled by ADS-B In, etc.) (Abbott & Rogers, 1993; Billings, 1991, Sarter & Woods, 1992, 1997; Woods & Sarter, 2000).

Abbott and Rogers (1993) highlight that there were good economic reasons to introduce a technology-centred solution, especially when the new design was expected to outperform the crew on a particular task or function. However, as Nakamura et al. (2013) point out, there are problems with this approach. The evolutionary addition of new systems has the same potential to change the role of the flight crew as a complete redesign would. There is a real danger that each new system added leads to increased crew workload, which in turn leads to reduced understanding of critical mission information. Furthermore, when the pilot's role moves from a system manager to a system monitor, entering and cross-checking data between systems without fully understanding the purpose of the activity, a loss of overall structured understanding is inevitable (Woods, 1996).

Automation was repeatedly sold as the preferred means to achieve a reduced pilot workload, however, studies found that the assumed benefit was in contrast with the overall systemic effect (Abbott & Rogers, 1993; Wiener, 1989). In simple terms, individual system design may not achieve a good solution when considered as part of the overall cockpit system, as that approach is not compatible with what we know about optimum coordination in human-machine systems (Bainbridge, 1983; Klein et al., 2004). A human-centred approach offers the key to being more successful. There is a fundamental difference between the goals of the traditional view, which tries to create a system that emulates human capabilities, and the non-traditional view of human-centred design, when the objective is to create a system that extends human capabilities and the main driver is to enable people by design solutions that matter for human purposes (Klein et al., 2004; Wickens et al., 2015).

As new flight systems technology was introduced, pilots became supervisory controllers, a role that focuses on monitoring and instructing lower-order automated subsystems (Wickens et al., 2015). This change also meant that a new form of cooperation and coordination emerged, as advanced automation was capable of independent action. Inagaki (2014) evaluated some of the issues that are at the centre of design solutions where humans and smart machines collaborate and cooperate sensibly in a situation-adaptive manner. He argued that the machine may be given authority to improve safety and to alleviate possible damage to the joint human–machine system, even in a framework of human-centred automation (Inagaki, 2014; Woods, 1996).

Inagaki (2005) also observes that in this new design context, the aim of automation is to work cooperatively with human operators in pursuit of shared objectives. To achieve that aim, function allocation needs to be dynamic and adaptive to appropriately support humans during their decision making process. When automation fails, current-generation airliners all rely on the human pilot as the ultimate fallback option (Dismukes et al., 2007). For that design assumption to be valid, pilots must always remain in command of their flight, and it follows that they must be actively involved in managing the system. Successful cooperation and coordination also assume that pilots and automated systems must understand each other’s intent in complex systems (Inagaki, 2005; Woods, 1996). For that to be effective, pilots must be adequately informed about the status and behaviour of automation through salient and understandable means (Billings, 1996; Parasuraman & Riley, 1997; Rankin et al., 2016; Woods, 1996).

2.5.3 Automation as a Team Player

Wiener and Curry (1980) raised concerns about whether total system safety was always enhanced by allocating functions to automated systems rather than human operators. Flight deck automation studies of B767 and B757 airplanes operated by major US airlines clearly indicated that most pilots preferred advanced technology over previous generations (Curry, 1985; Wiener, 1989). A decade later, once newer generation airliners became the norm, attention turned to exploring the issue of designing flight deck procedures for automated cockpits (Degani & Wiener, 1997). Aviation safety advocates were looking for improved guidance to pilots, ensuring safe, logical, and predictable (standardized) means of carrying out flying duties on the line (Wiener & Nagel, 1988).

Based on a comprehensive examination of common ground and coordination principles, and the fundamentals of human-centred design, Klein et al. (2004, 2005) observe that the real challenge is to make automation a team player in joint human-machine activity. As part of maintaining common ground (during coordinated activity) team members must direct each other's attention to the most important signals. In this context, *common ground* refers to the sum of the agents' common beliefs and knowledge, or the mental model of the system as a whole (Klein et al., 2005). The coordination, however, must be achieved in an intelligent and context-sensitive manner, not to overwhelm each other with low-level messages (Brennan, 1998; Clark, 1996).

In the aviation domain, research studies clearly show the impact of automation on human cognition and behaviour, especially when new complexities and new forms of error are introduced (Billings, 1991; Sheridan, 1992). Automation often fails to function as a team player (Klein et al., 2005). Automation that is strong, silent (when no or poor feedback is provided about actions taken and future intentions), clumsy (when it interrupts humans during high workload or high tempo periods), or is difficult to interact with, does not support a coordinated joint effort (Norman, 1990; Sarter, 1994; Sarter & Woods, 1997). When communication and collaboration break down, it almost always leads to systemic failure and potential safety loss (Eriksson & Stanton, 2015).

Modern airliners face the same challenge. Woods (1996) highlights that automation can compensate for the transition of control, but often does that silently. When the crew is unaware of trouble until automation nears its limit of authority or capacity to compensate, the crew takes control too late, or may not be ready to take control and recover the airplane from an undesired state. It is crucial that human-machine collaboration provides contextual and temporally relevant information communicated in a safe and reliable manner. This ensures that the mental model of humans remains up-to-date and can cope with sudden transfer of control. This is a classic challenge of highly automated interactions (Klein et al., 2004).

Eriksson and Stanton (2015) offer the AF447 accident as an example when the pilot is expected to take over while suddenly going from a *mental underload* (standby or low workload tasks) to a *mental overload* (which requires operating at a peak capacity). Woods (2021) evaluates more recent examples of 737 MAX family accidents when misbehaving automation (sensor failures feeding incorrect data to a powerful new automated system) acted on its own, without human input, and there was no feedback about the actions and future behaviour of the automated system.

When automation is not a team player, it leads to new problems and new forms of system failures (Woods, 1996). Perrow (2011) highlights that high coupling leads to new forms of breakdowns. Consequently, there can be a cascading effect when individually small disturbances link up, and failure symptoms may appear in seemingly unrelated parts of the system (Reason, 1990). Interaction-type systemic failures make fault management and diagnosis much more complicated (Perrow, 2011). This effect can have a critical bearing on crew members' in-flight cognitive workload during a check flight.

Sheridan (1992) finds that when an automated system design is based on a static view of allocating functions, it leads to failure. The first issue to overcome is that automated agents should coordinate actions with other agents. Or realistically, the human agents should at least be able to see the actions and intentions of automation so that humans can coordinate via other autonomous (subordinate) agents. The second issue to be addressed is the need to calibrate the human operator's over-trust in and over-reliance on automation (Billings, 1991). Inagaki (2014) argues that different types of over-trust and over-reliance may vary depending on the characteristics of the automated system.

Woods and Sarter (2000) suggest that if the behaviour is not salient and understood, and coordination breaks down between the human and machine agents, such as the pilot(s) and a highly automated airliner, there is a high likelihood that the pilots' expectations about automation behaviour are not met. Automation breakdowns can manifest in unexpected action, the failure to take expected action, or an action carried out in an unexpected manner. Sarter and Woods' (1997) earlier research study of a popular fourth-generation jet airliner showed that most line pilots (80 percent) reported that they had experienced automation surprises at least once. A similar survey of a competitor's third-generation model found that 2/3rd (67 percent) of the responding pilots confirmed that automated systems sometimes behaved in unexpected ways (Javaux, 1998; Sarter & Woods, 1992). The 1972 L-1011 Everglades accident was a textbook example of such a *fail silent* problem and unforeseen limitations. The NTSB (1973) found that the crew was preoccupied with trouble-shooting a landing gear indication light, while the altitude hold function of the autopilot silently disconnected, a poor feedback on the automation state. The disengagement was not salient, an inherent design issue (Norman, 1990; Sarter & Woods, 1997).

Inagaki (2005) observes that the 1994 A330 test flight crash in Toulouse can also be explained as an automation surprise event, this time between the pilots and the modified A330 design. Based on evidence uncovered during the investigation, it appears that the crew was surprised when the automation performed tasks which were not explicitly directed by the pilots. The preliminary report cited evidence which indicated that the crew was confused about the airplane's behaviour in a critical flight phase (DGA, 1994). Funk et al. (1999) cite the 1996 DC-8 check flight accident at Narrows as an example of an automation-related omission error. Notable accidents from routine line operations were the 1994 A300 accident at Nagoya where the likely immediate cause was an inadvertent activation of the autopilot go-around mode (Aircraft and Railway Accidents Investigation Commission [ARAIC], 1996), and the 1995 B757 accident at Cali where the investigation report cited FMS over-reliance (NTSB, 1996).

As a result of the Nagoya and Cali accidents, the FAA launched a landmark human factors study to identify any specific or generic problems in design, training, flight crew qualifications, and operations, and to recommend appropriate means to address these problems (Abbott et al., 1996). Abbott et al. (1996) identify issues that show vulnerabilities in flight crew management of automation and situation awareness, including pilot's understanding of automated systems, automation modes, capabilities, and limitations, which can lead to automation surprises and mode awareness issues. The FAA study finds variability between pilot decisions about the appropriate automation level to use, especially in non-normal situations, and potential mismatches with the designer's assumptions about how the pilots will use automation. The study also highlighted concerns about flight path awareness, including insufficient terrain awareness, which can lead to Loss of Control – In Flight (LOC-I) or CFIT, and concerns about energy awareness (especially low energy states) (Abbott et al., 1996). This landmark FAA study largely contributed to airlines establishing *automation policies* in support of their line operations (Wiener et al., 1999).

Aviation accident reports do not support the theory of a pilot being out-of-the-loop or described as a complacent supervisor of automated systems who somehow loses situational awareness (Billings, 1991; Endsley & Kiris, 1995; Wiener & Curry, 1980). It is rather the irony of coordination when strong and silent automation leads to airline accidents (Sarter & Woods, 1997). The aircraft can easily be managed into a disaster when the pilot is an active manager and coordination breaks down between the pilot and the highly automated machine (Harris, 2003; Parasuraman et al., 2000). Automation is pure logic, and it does exactly what it was programmed to do (Dekker & Hollnagel, 1999; Dekker & Woods, 2002).

Harris (2003) observes that the pilot's interaction with automated systems is not a unidimensional concept. Coordination breakdowns are not necessarily interpreted the same way by a pilot, a design engineer, or an accident investigator. These different perspectives, and the underlying assumptions about what is the right choice in an unexpected scenario, all need to be considered when evaluating causal explanations for accident scenarios (Leveson, 1995, 2014), as follows:

From a design engineer’s view, automated systems behave in a deterministic fashion (Leveson, 1995, 2014). Woods (1996) observes that, in theory, if one has complete knowledge about system rules, has access to all past instructions and has total awareness of the external environment, then that person would be able to accurately project automation behaviour. In practice, the cognitive demand of projecting the behaviour for highly complex systems is very significant (Woods & Hollnagel, 1987). For example, Veillette’s (2009) knowledgeable technical pilot argument would require such a reasonably accurate projection during an unexpected check flight scenario. Without robust feedback about the current and future state and behaviour of automated systems, it is unreasonable to expect that the pilot can reach a correct conclusion.

From an accident investigator’s view, there is a fundamental difference between what the operator experiences in context and what the investigator sees in hindsight, i.e. for the pilot the system does surprising things on its own, while the external observer can find the same behaviour as a direct and rational result of past instructions and current state of automated systems (Dekker, 2002; Woods, 1996). This safety paradox has strong implications for causal analysis and accident reconstruction for complex automated systems, like modern airliners. Current industry trends reflect automated systems with increasing authority and changing feedback mechanisms, which only reinforce strong and silent automation characteristics and latent dangers of automation (Billings, 1991; Norman, 1990; Woods et al., 1994).

In-service experience with modern (third and fourth) generation airliners confirms that line pilots can successfully negotiate common system and component malfunctions during normal operations (Airbus, 2022; Boeing, 2022). At the same time, insufficient system knowledge or understanding of actual and future aircraft states may decrease pilots' ability to respond to failure situations. This is a particular concern for failure scenarios which do not have procedures or checklists, or where the procedures do not offer a recovery path (Nakamura et al., 2013).

On a more positive note, Dismukes et al. (2007) find that the systemic incorporation of human factors principles in automated flight deck designs over the last 50 years enabled a potential solution in reducing automation-related accident rates. When pilots are provided more explicit information through better displaying the information, supplemented by conservative decision making guidance in ambiguous situations, the general mismatch between automated cockpits and human cognitive characteristics becomes less powerful in misleading the flight crew.

2.5.4 Section Summary

In this section, the literature review was broadened in scope to explore the evolution and in-service experience of highly automated flight deck systems, looking for potential systemic factors which may contribute to non-routine accident scenarios.

A review of flight systems development found evidence that early (first and second) generation designs were supported by an incremental approach, primarily driven by the availability of new technology. In this technology-centred approach, the designer's focus was on higher levels of automation, the machine progressively taking over functions previously carried out by the pilot. When more complex systems were introduced to current (third and fourth) generation airliners, the distance between the pilot and direct control of the airplane grew even further, which led to the pilots understanding even less about actual automation states and future behaviours.

In a human-centred design context, the aim of automation is to work with the human operator in pursuit of shared objectives. Despite significant advances made over the years, when automation fails, current-generation airliners continue to rely on the human pilot for initiating a recovery attempt. In the aviation domain, automation often fails to function as a team player. When communication and coordination break down between a highly automated airliner and the flight crew, it almost always leads to systemic failure and potential safety loss.

Check flying is likely more sensitive to a sudden transfer of control than routine flight operations. In this critical and high-tempo environment, it is even more important to communicate contextual and timely information about the automation's state and future behaviour and to do so in a safe and reliable manner. A potential safety solution is offered by providing the crew with more explicit and better communicated information, supplemented by conservative decision making guidance in ambiguous situations.

The next section revisits the literature to understand the latter part, exploring research studies about human decision making and the evolving concept of human error in uncertain and time-critical settings.

2.6 Human Decision Making

Maintenance test flights represent a complex real-world setting. When things do not go according to plan, the crew is expected to make critical decisions in a high-tempo and uncertain environment. Following on from unwanted effects associated with automated systems, this section reviews the literature to understand what new error traps were introduced in the flight deck environment and what can be done to support the crew’s decision making process.

2.6.1 Decision Theory

Howard (1988) establishes a central distinction in the field of decision analysis between normative and descriptive views of decision making. Decision making in a traditional sense is about selecting from suitable alternatives. While decision analysis is supported by hundreds of years of philosophical thought about uncertainty and decision problems, the roots of modern decision theory can be traced back to the formative work of Edwards (1954), which introduced a formal normative view and the concept of humans’ inability to act as rational decision makers (Edwards, 1954, as cited in Hollnagel & Woods, 2005, p. 194).

The discipline of applied decision analysis was founded on normative theories of decisions, which explored decisions in the context of real-world behavioural economics (Howard & Matheson, 2005; Keeney, 1982; Raiffa, 1968). Early research examples range from purely mathematical (Von Neumann & Morgenstern, 1947), through multiple objectives and subjective expected utility theory (Slovic et al., 1977), to statistical models (Edwards et al., 2007). Further advances were made when the theory was extended with additional prospective and behavioural elements or more recent approaches based on Bayesian network models (Cox Jr, 2015).

Over the years, decision analysis techniques and models were applied to various engineering, supply chain, and safety risk problems, however had a limited impact on real-world decisions, primarily due to limits posed by simulation models and the inherent characteristics of locally rational decisions, which converge the model on a suboptimal outcome (Carriger & Barron, 2011; Howard, 1988; Howard & Matheson, 2005; McNamee & Celona, 2007). When modified with game theory elements, the theory was more successful in modelling risk in complex systems, like electrical network grids (Cox Jr, 2015; McNamee & Celona, 2007).

The literature refers to two major milestones in modern decision theory: the principle of approximate decisions, more commonly known as Simon's (1955) concept of satisficing; and the introduction of Natural Decision Making (NDM) theory. NDM models frame decisions as a sensemaking problem, rather than a distinct decision making process (Klein, Orasanu, Calderwood, & Zsambok, 1993; Orasanu & Connolly, 1993). The following sections provide a summary of new approaches to decision theory in the context of aviation decision making and decision errors contributing to safety outcomes.

2.6.2 Naturalistic Decision Making

In contrast to earlier decision theories, NDM theory describes how human decision makers *actually make decisions* in complex real-world settings, rather than how they were supposed to make the right decision. NDM's focus has always been on situating human decision makers in complex systems and the engineering and technological aspects of highly automated systems and their influence on the decision making process remained secondary (Cook, 2012). The origins of NDM theory go back to the 1980s, when Orasanu established an army research program, which led to the discovery of the Recognition-primed Decision (RPD) model by Klein and his colleagues (Klein et al., 1986). Since then, several other NDM models have been developed by the research community, although RPD remains the most widely known and adopted NDM framework (Klein et al., 1993; Klein & Wright, 2016; Orasanu & Connolly, 1993; Orasanu & Fischer, 1997).

Klein et al. (1986) describe the formative idea of RPD as experts recognising patterns in real-life situations they encounter. Recognition refers to a limited set of options available to the expert. The model does not have any traditional (normative) elements, like calculations or estimates. Experts arrive at the decision quickly by reaching back to their memory and past experience. RPD describes different levels of complexity which involve a combination of situation assessments and mental simulation, as demanded by the decision context. The theory is robust and well-researched in critical operational domains, but the supporting concept of situation awareness remains less clear-cut, which will be explored in further detail below (Cook, 2012; Shattuck & Miller, 2006).

According to the NDM doctrine, eight important factors characterise decision making in naturalistic settings (Klein et al., 1993; Orasanu & Fischer, 1997). These factors echo the decision environment faced by a crew during any check flight scenario, as follows:

1. Ill-structured problem – even if test points are clear and understood, the decision problem does not come in a complete form. There is considerable cognitive effort involved in assessing the in-flight situation or even recognizing that a decision is required.
2. Uncertain dynamic environment – NDM describes an environment where information may be ambiguous or incomplete or simply poor quality, further aggravated by a rapidly changing environment.
3. Shifting or competing goals – in a check flight context, the classic trade-off is between time, test objectives, and safety margins.
4. Action and feedback loops – normative theories refer to a single decision event at a point in time. NDM is about a whole series of decision events, either dealing with the problem or exploring it.
5. Time stress – the decision maker is under significant stress. The relevant timescale for a check pilot is from minutes to seconds, or

even milliseconds, as the situation escalates. NDM recognises the shift to less complicated reasoning, as personal stress levels rise.

6. High stakes – post-maintenance test flights must safeguard lives (crew members and people on the ground) and preserve high-value assets, like the airplane and critical infrastructure.
7. Multiple players – such as flight and ground crew members, ATC, management, customers, etc.
8. Organisational goals and norms – values and goals may not match the personal preferences of crew members. Post-incidents, it is common for operators and regulators to establish new SOPs, rules, or decision making guidelines.

Orasanu and Martin (1998) extend the NDM framework to study errors in aeronautical decision making (ADM) as a factor in accidents and incidents. Instead of slips and memory lapses, their research focused on errors of intention where the source of error is either in knowledge gaps or in the process of reaching the decision (Norman, 1986). Realizing that incident outcomes (the safety loss) are only loosely coupled with the actual decision event, they looked at the process by which decisions are made. That resulted in framing the decision process as the combination of situation awareness and choosing the right course of action (Orasanu & Fischer, 1997). It follows then that decision errors either originate from a wrong interpretation of the problem or from choosing the wrong course of action (Orasanu, Martin, & Davison, 1998; Woods, 1996).

An NTSB (1994) study evaluated a set of 37 accidents which were originally attributed to the flight crew as some form of tactical decision error. After revisiting those accidents, Orasanu and Martin (1998, pp. 102-103) put forward the proposition that pilot tactical errors can be traced to a) ambiguity of cues, which results in the situation not recognised, and in turn, the need to change course is not triggered; b) the likelihood and consequences of risk not perceived or underestimated by the crew; and c) conflicting goals and values, such as the classic airline trade-off between safety and schedule, or other social factors.

Orasanu and Martin's (1998) ADM framework, as a subset of NDM theory, led to rediscovering that pilots operating modern airliners need to be better supported during the decision making process. In ADM terms, improved situation awareness would require better and more accessible diagnostic information, integrated displays, and predictive models that inform the pilot about risk levels and remaining time available (Wiener et al., 1999). Course of action errors would be mitigated by improved simulator training for worst-case scenarios and training in what-if reasoning (Orasanu & Martin, 1998). In theory, a decision aid which would present context-sensitive risk estimates for the pilot would be a major step forward in supporting the right course of action (Burdun, 1998; Burdun & Parfentyev, 1999). Preferably the information would be presented in a graphical format, but there is no published evidence that such a concept had been successfully integrated and validated in a complex decision making environment, such as a modern flight deck.

Mosier (2008) makes the case that sophisticated technology and modern systems have fundamentally altered the naturalistic work environment, a profound change which resulted in a hybrid ecology, a combination of discrete technical space and the continuous ecology of the natural world around us. She posited that this profound change also impacted the critical macrocognitive processes (sensemaking, coordination, and dealing with ambiguity) underpinning NDM theory (Klein et al., 1993). In a check flying context, Mosier's (2008) most critical finding is the realization that expert intuition, which is the foundation of the NDM framework, may not be effective in hybrid environments or may even create a hazard to the decision making process. When testing the correct functioning of automated systems, automation-induced omission errors (when pilots do not take appropriate action) or commission errors (when pilots blindly follow automated directives) may occur when the pilot heavily relies on automated cues (Mosier et al., 1998).

Mosier (2008) also challenges the notion of traditional human-focused NDM research and provides a refined understanding for this research project about the powerful impact technology had on the decision making environment in a modern airliner’s cockpit:

- A highly automated digital flight deck has a significantly different NDM environment when compared to early-generation airliners and their analogue instrumentation.
- A pilot is an expert decision maker, but expert intuition is not necessarily effective and sufficient in the new hybrid environment. Making sense of data presented in a digital cockpit poses a new challenge.
- Classic NDM approaches do not allow for analysis during decision making. In a highly automated digital cockpit, critical macrocognitive functions cannot be accomplished without analysis.
- The designer’s assumption that intuitive displays eliminate the need for the pilot to analyse the situation is not a valid approach. Modern systems cannot be managed intuitively, as highlighted by events like the 1992 A320 accident at Mont Sainte Odile (BEA, 1993).
- Technology in this new hybrid environment does not necessarily support human cognition. Digital flight deck systems layer and mask data before presenting it to the crew, considered appropriate by the designer and the certification process, forcing the pilot to locate system data needed for analysis.

In summary, an opaque cockpit interface, combined with the sequence and layered presentation of digital data, can easily short-circuit the macrocognitive processes. In this context, even an expert decision maker needs to focus on the most salient cues, which activates internal heuristics (Billings, 1996; Mosier, 2008; Mosier et al., 1998; Woods, 1996).

2.6.3 Sensemaking

In 1979, the nuclear reactor at Three Mile Island suffered a major accident when the reactor core was exposed and melted (Perrow, 1981; Rogovin, 1980). It was a watershed moment in nuclear safety, which led to substantial research activity over the years, including the role of human operators and their mental model of the world and how it works (Norman, 1986). Before 1990, the focus was on how to generate the right symbols to generate the correct mental model for operators (Rasmussen, 1987). From 1990 onwards, research into sensemaking by operators was a follow-on from those earlier studies (Cook, 2012).

Weick et al. (2005) observe that sensemaking involves integrating new information into an existing model through the process of comprehension and development of plausible images in retrospect. Through this organizing process, people can extract cues, construct explanations, or even make predictions. In everyday life, when people are confronted by an event they do not fully understand, sensemaking tries to answer the questions of what is going on and what should be done about it. Those two questions bring events and their meaning into focus, hoping that by finding answers, they can remain in touch with what is happening around them (Woods, 1996). The notion of sensemaking can be traced to organisational research, however, the concept soon became a study subject for decision making. Contrary to decision analysis, though, sensemaking is not concerned with decisions made by an individual, its sole focus is on context and interpretation (Klein et al., 2006a). Framing a problem as a question of meaning rather than decision making can help mitigate the risk of fundamental attribution bias (Heider, 1958). To borrow Snook's (2000, p. 178) excellent insight, accidents should be seen as "good people struggling to make sense", not as "bad ones making poor decisions".

Weick's (1988) seminal article is one of the first examples in the literature that examines sensemaking in the context of a major accident, drawing on Shrivastava's (1987) excellent analysis of the Bhopal chemical disaster. Maitlis and Sonenshein (2010) observe that retrospective sensemaking studies of crisis events fall into two main streams. Research studies can either describe how sensemaking unfolds during a crisis or address how sense is made after a disaster, drawing on public safety inquiries. Notable examples in the first group include the Bhopal industrial accident (Shrivastava et al., 1988; Weick, 1988, 2010), the Tenerife air disaster (Weick, 1990), the Space Shuttle Challenger explosion (Vaughan, 1990, 1996), or the collapse of sensemaking at Mann Gulch (Weick, 1993). Gephart’s (1993) post-inquiry study of a fatal pipeline explosion offers an ethnomethodological example from the second group. His constructivist approach describes an integrated qualitative framework that involves theoretical sampling (Glaser & Strauss, 1967), data and textual analysis, and expansion analysis (Gephart & Saylors, 2020) to generate insights into organisational sensemaking about operational hazards and risk management, and the way a public inquiry apportions blame.

Klein et al. (2006b) offer Data-Frame Theory as an empirically grounded account of sensemaking. In a sensemaking context, frames refer to previously learned mental structures that guide people when processing and interpreting new information. Conceptual models of pilot perception and actions regularly use framing and reframing metaphors when describing and explaining pilot performance. Reframing refers to the process when a person is trying to make sense of what is happening around them, following an unexpected disturbance, a surprise. Framing is an adaptive process, as frames not only define what is meaningful data, they also change as relevant data is obtained (Rankin et al., 2016).

Research studies into dealing with unexpected events in the cockpit are directly relevant to the check flight safety problem. Recent examples adapted the reframing metaphor to startle and surprise events (Landman et al., 2017) and a crew-aircraft sensemaking model (Rankin et al., 2016) for improving pilot responses to unexpected events. The reframing studies found that sensemaking is vulnerable to startle effects and acute stress. Interestingly, both studies concluded that safety interventions should primarily focus on improving pilot skills and control strategies, a finding that is not fully compatible with optimum coordination in human-machine systems (Woods, 1996). Oliver et al. (2017) applied limits concepts to the Air France AF447 accident, including limits to cognition, which is a fundamental constraint in the sensemaking process. The AF447 analysis highlighted what can happen when automation limits suddenly disappear. Once flight envelope protections were removed, the pilots could not comprehend what was going on, exposing their cognition-based limits.

2.6.4 Cognitive Systems Engineering

Hollnagel and Woods (1983) observe that as automation levels increased in commercial airliners, the pilot’s role changed to a monitoring and supervisory capacity, which in turn, shifted the demand from motor skills to cognitive decision making skills. The operator uses a mental model of the automation (machine), and the machine has a pre-defined image of the operator. The designer’s challenge is to match the two agents on a cognitive level, not only in a physical sense. With the increasing sophistication of software and hardware systems, the human-machine interface has become an intersection of two cognitive systems. In that context, a cognitive system can be defined as an adaptive system that uses information about both the environment and itself when planning and updating its actions.

Hollnagel and Woods (2005) describe Cognitive Systems Engineering (CSE), an approach based on Bertalanffy's (1969) general systems theory. CSE deals with how joint cognitive systems cope with complexity, especially dynamic aspects. Essentially, CSE is an attempt to describe how humans cope with the complexity and unpredictability of their environment. A major field of application for CSE is supporting decision making. Traditional decision making is about choosing what to do where decision making is a distinct process. NDM sees decisions as a sensemaking process. CSE deals with the ways to carry out a chosen alternative in a complex world, not the choice itself. Not surprisingly, CSE shifts the question from overcoming limits to supporting adaptability and control (Hollnagel & Woods, 1983, 2005)

Woods and Hollnagel (2006) identify the joint cognitive system as the base unit of CSE studies, not the human operator or the machine in isolation. The target of the investigation is to discover patterns in cognition which can then be used as design concepts and techniques for system development. Patterns here refer to reoccurring problems in the system's environment and the design solution to that problem

(Alexander, 1977; Woods, 2021). What emerges from a CSE study are patterns in coordinated activity, in resilience (the ability to adapt to surprise and error), and in affordance (how artefacts support people's natural ability and expertise) (Hollnagel et al., 2006; Woods & Hollnagel, 2006). Hollnagel et al. (2006) also highlight that in a CSE context, when pilots control an airliner, they need to do more than blindly follow rules and SOPs. A pilot needs to anticipate hazards, search for potential conflicts, cope with surprise, accommodate changes in the operating environment, come up with suitable workarounds, resolve the difference between plans and the actual situation faced by the system, and recover from incorrect assessments of the situation or communication errors.

More recently, Roth's (2018) transactional approach extends CSE theory and joint cognitive systems for the purposes of forensic cognitive science. A transactional approach focuses on the internal cognitive dynamics of the joint cognitive system, looking for reasons why the whole system acted in a particular way, as revealed by post-accident analysis (Woods & Hollnagel, 2006). The method was successfully applied to retrospective accident analysis and confirmed Dekker's (2004) earlier findings about shortfalls in investigations that focus on traditional human error and Endsley's (1995) situation awareness concepts (Roth, 2018).

2.6.5 Human Error

A large portion of accident causes are attributed to human error, without understanding why it made sense for people to do what they did (Baker et al., 2008; Shappell et al., 2007). Cognitive biases can influence accident investigators’ interpretation and reasoning about causal factors and traditional investigation methods fail to prevent their negative influence (Dekker, 2002; Dismukes et al., 2007). Attribution bias (Heider, 1958), hindsight bias (Fischhoff, 1975), and confirmation bias (Nickerson, 1998) are very powerful examples. This section reviews the prevailing human error paradigm in the context of causal explanations, including the cognitive biases mentioned previously, attempts to reduce bias in causal judgements, and new approaches and ideas promoted by safety researchers.

Heider’s (1958) attribution bias refers to the systemic error made when people are trying to find reasons for their own or others’ behaviour. People continually make causal attributions; however, those attributions are not always accurate which results in a biased interpretation. The concept originated as a field of inquiry in the 1950s, and later expanded to identify relevant conditions, fundamental types of attribution error, and the inherent actor-observer bias (Kanouse et al., 1987). The actor-observer asymmetry partly explains why investigators (observers) and pilots and engineers (actors) may provide different causal explanations for check flight events. Actors tend to focus on what is special about the situation that caused their behaviour. Observers treat the situation as background and focus on what is special about the actors that makes them different from the rest of their population, in this instance the airline pilot and engineering community (Adler, 2008). Interestingly, attribution theory describes humans as very effective processors of information, which is in sharp contrast to judgement research. Attribution researchers make explicit assumptions about people's ability to identify causal relations, while judgement research is looking for more and more biases in people's judgements and decision making (Fischhoff, 1975; Kanouse et al., 1987).

Fischhoff (1975, 1982) describes hindsight bias as the tendency to assume that one knew that a given outcome is likely. His research studies evaluate the major difference between historical and non-historical judgment where the historical judge typically knows how things turned out. In his experiments, knowledge of the outcome was found to have increased the predicted likelihood of reported events and to have changed the perceived relevance of event-descriptive data, regardless of the likelihood of the outcome. The judges are largely unaware of their bias and this lack of awareness can seriously restrict one's ability to judge or learn from the past (Fischhoff & Beyth, 1975). Given knowledge of the outcome, accident investigators tend to simplify the problem that was faced by the incident crew. Decisions and actions that led to a safety loss are judged more harshly than if the same decisions had resulted in a neutral or positive safety outcome. Unfortunately, hindsight bias is present in many investigation reports even when investigators have been warned about the effect (Woods et al., 1994).

Confirmation bias is a tendency to seek out evidence which confirms rather than contradicts current beliefs (Nickerson, 1998). The literature on reasoning suggests that when people encounter arguments which support existing beliefs, they tend not to examine the evidence closely. The situation when confronted with evidence that contradicts one's beliefs is different. To accept such an argument without scrutiny would disrupt the belief system (Evans et al., 1993). Sloman (2008) found that when explanations compete, the credibility of a statement is proportional to the success of the attempt to create an explanation that is consistent with one's knowledge and beliefs. Research looking at experts has found the opposite effect. Klein et al. (2006b) cognitive task analyses found that once expert decision makers detect that their current frame is inaccurate, they shift to seeking a competing frame, deliberately looking for information that disconfirms their existing hypothesis. What might look like confirmation bias is simply using frames as a guide. In terms of accident investigations, discounting competing explanations may be the result of the investigator's desire to provide a coherent package. The act of constructing a causal explanation may cause the investigator to neglect other alternatives (Sloman, 2008).

By recognizing the impact of cognitive biases in retrospective accident investigations, it can be seen why it is so tempting for investigators to accept the known outcome as inevitable, which simplifies the task of pointing out instances of human error. This approach, however, does not help prevent recurrence. Dekker (2002, 2003, 2004) proposes a new approach when investigating human contributions to accidents, by focusing on actions in a new light, to try and understand why assessments and actions made sense to people at the time. There are complex interactions between the pilot, the task demands, the environment, organisational and social factors, etc. Accident crews respond to the unfolding situation as they perceive it, bounded by local rationality. Dismukes et al. (2007) find that experts do what seems reasonable, given what they know at that moment. Experts making an error is not de facto evidence of missing skills or negligence. The main reason pilots remain in a highly automated cockpit is the superior capability of the human brain to come up with the right decision in ambiguous and uncertain situations (Dekker & Woods, 2002).

Amid the prevailing human error paradigm, researchers started to advocate a powerful alternative in the early 1990s. Rasmussen (1990) observes that human error cannot be removed from complex systems by improved system design or better training. He concludes that operators should rather be supported by means of recovery when an inevitable error occurrs. In a similar vein, Van der Schaaf (1995) promotes the idea of *human recovery*. He posits that by giving a new context to near misses (incidents), by tracing the positive role that humans play in preventing small errors from cascading into a major failure, the real causes of recovery can be revealed. The updated human event analysis guidelines adopted by the United States Nuclear Regulatory Commission (NRC) offer an example. A Technique for Human Event Analysis is a retrospective analysis process designed to identify human failure events and their underlying causes in nuclear incidents (NRC, 2000). In essence, the updated guidelines describe a second-generation human reliability analysis process for probabilistic safety assessments. The NRC (2000, p. 145) postulates that unsafe human action occurs in an error-forcing context that can be identified during an investigation. The key outcome of the NRC (2000, p. 149) analysis is to identify the causes of unsafe acts and to describe the performance-shaping factors and significant plant contributing factors.

Similarly, Dismukes et al. (2007) offer a new approach to uncovering the types of error to which expert commercial pilots are vulnerable and to identify the underlying cognitive, task, and organisational factors. Their case studies revisit 19 historical NTSB reports which primarily attributed probable causes to pilot error. The starting point for each case study is the NTSB's operational and human performance analysis of the accident. In contrast to the NTSB's mandate, they attempt to understand the nature of vulnerability of all similarly skilled pilots to error in situations described in the relevant investigation report. A central theme emerging from their review is the difference between multiple factors influencing the probability of making random errors and investigators citing such factors as causal to pilot errors and, by extension, to the accident outcome.

2.6.6 Section Summary

This section of the literature review explored human decision making in the context of non-routine flying, especially the additional challenges created by operating in a high-tempo and uncertain environment. The review focused on the evolution of decision theory, from traditional decision analysis, through NDM theory which sees decisions as a sensemaking process, to relevant aspects of CSE theory and joint cognitive systems, followed by revisiting new approaches to investigating human contributions to accidents.

Contrary to earlier decision analysis techniques, NDM theory accurately describes the complex real-world decision making context of check flight operations (Klein et al., 1993; Orasanu & Fischer, 1997). Mosier (2008) extends NDM research by evaluating the impact of modern digital technology on the decision environment. In a check flying context, her most critical finding is the realization that expert intuition, which is the foundation of the NDM framework, may not be effective in hybrid environments or may even create a hazard to the decision making process.

The notion of sensemaking is central to the NDM doctrine. Weick (1988, 2010) describes the difficulty involved in adequate sensemaking in crisis situations and the risk involved in humans unwittingly escalating the crisis in technically complex and highly interactive environments, especially when facing non-routine problems. Check flight incidents present such low probability and high consequence events where judgement can rapidly deteriorate under pressure. Weick (1988, 2010) posits that changes enacted on the environment are internalized by human actors as plausible causal maps by which observed actions produce observed consequences. The enacted environment is also visible to an external observer who can infer traces of cognition and causal relationships. Studying investigation reports offers such an avenue to learn more about sensemaking in crisis situations, including events triggered by human actions that may unfold into serious incidents or accidents.

At the same time, the accident analyst needs to be careful not to limit the sensemaking study to a limited timeframe, as that approach may lead to an oversimplified accident model that only reveals a few organisational factors (Weick, 2010). While recent studies delivered new insights into surprise and startle events in the cockpit, the reframing approach is necessarily limited to compressed timeframes (Klein et al., 2006b; Landman et al., 2017; Rankin et al., 2016). A narrow analysis window is not suitable for uncovering a complete set of systemic causes across multiple check flight occurrences.

Finally, the review identified a powerful alternative to the prevailing human error paradigm. By tracing the positive role human pilots play in preventing cascading errors in complex systems and looking at human contributions to accidents from the angle of why their decisions made sense to them at the time, common biases can be mitigated when attributing causal factors. In terms of retrospective accident analysis, comparing traces of human recovery attempts in incident and accident reports can offer new insights into systemic causal factors. Dismukes et al.'s (2007) decision to change the retrospective analysis from looking for probable causes to evaluating the reasons why other expert crews would be vulnerable in a similar in-flight scenario offers a template for revisiting pilot error findings in check flight investigation reports.

2.7 Air Safety Investigations

The final section of the literature review explores the theoretical foundations and practical implications of air safety investigations. It is beyond the scope of this research project to revisit the differences between ICAO Contracting States' individual legal and investigative frameworks. The aim is to describe the accident theories behind recommended practices, problems with accident causation models, and new approaches in developing accident models that are suitable for complex systems, such as a highly automated airliner.

2.7.1 Accident Investigation Standards

The Chicago Convention imposes an obligation on the State of Occurrence to investigate accidents and incidents in accordance with ICAO International Standards and Recommended Practices (SARPs) (ICAO, 2020a). ICAO SARPs laid out in Annex 13 to the Convention are universally adopted as the *de facto standard* of accident investigations, albeit applied with differences between member states. ICAO (2020a, p. 25) declares that "the sole objective of the investigation of an accident or incident shall be the prevention of accidents and incidents. It is not the purpose of this activity to apportion blame or liability.".

A range of ICAO documents provides information and guidance on subjects related to accident investigations. This section is limited to reviewing the overall framework and the theoretical accident models underpinning the recommended standards.

The ICAO Framework. The ICAO (2011) investigation process seeks to identify causes by a process of elimination, whereby all possible causes of an event are evaluated and eliminated until the remaining causes are identified and recorded in a standard format investigation report. In the overall safety management context, ICAO (2018, 2020a, 2020b) defines the role of accident and incident investigations as a reactive component, which contributes to improving the aviation system by identifying the root causes of accidents and lessons learned from those events. In a system safety context, the ICAO (2020b) recommended objective is to identify all immediate and underlying systemic causes, supported by safety recommendations aimed at hazards and eliminating deficiencies.

Within the ICAO (2020a, 2020b) framework, causes refer to events which resulted in a safety loss. A cause can be an act, omission, condition, or circumstance which should be eliminated or avoided to break the causal chain. This is the world through the eyes of Reason (1990). The standard for the causal analysis process is to present any condition, act, or circumstance in a chronological fashion. In other words, the ICAO standard is a *root cause analysis* (RCA) process. Safety boards do not necessarily follow the RCA standard, for the reasons explained in the next section of the review. For example, the Australian Transport Safety Bureau (ATSB), the AAIB, or the Transportation Safety Board of Canada

(TSBC) do not identify *probable* or *primary* causes for accidents (Canadian Transportation Accident Investigation and Safety Board Act, 1989; Civil Aviation (Investigation of Air Accidents and Incidents) Regulations (UK), 2018; Transport Safety Investigation Act, 2003). In contrast, the NTSB is charged by the US Congress to find the *probable cause* of accidents, a textbook application of the RCA method (Independent Safety Board Act of 1974, 2006).

Reference Investigation Methods. ICAO (2011) adopted several sequential and epidemiological accident models in its guidance. Sequential models describe a linear chain of cause-effect relations, while epidemiological models are an analogy to a spreading disease, a combination of active and latent factors (Hollnagel, 2002). Reason's (1990) model, the Management Oversight and Risk Tree (MORT), the Human Factors Analysis and Classification System (HFACS) and the Threat and Error Management framework (TEM) form the primary ICAO toolkit (as cited in Helmreich et al., 1999; Hollnagel, 2002; and Sklet, 2004). HFACS is an extension of Reason's model (Wiegmann & Shappell, 2000). While ICAO (2011) considers these models to be sufficient systemic models, it also encourages Contracting States not to limit their toolkit to the accident models listed in the reference framework. The literature contains multiple studies which compare selected methods for accident analysis (Herrera & Woltjer, 2010; Leveson, 2001; Sklet, 2002, 2004).

ICAO's (1993) guidance for human factors investigations establishes human element objectives, which largely reflect Hawkins and Orlady’s (1993) view of human performance and contribution to occurrences. The ICAO (2011) recommended integrated process relies on a combination of the modified SHEL(L) model (the name derived from its components, software, hardware, environment, and liveware), Reason’s (2008) Latent Unsafe Conditions and Generic Error-modelling System frameworks. The SHEL model was originally developed by Edwards (1972) and later modified by Hawkins (1975) to become known as SHEL(L) (Hawkins & Orlady, 1993). A SHEL(L) investigation focuses on components in the model, especially the interactions and mismatches between liveware, software, and hardware elements (ICAO, 2018; Wiegmann & Shappell, 2001).

2.7.2 Accident Causation

Evaluating large-scale accidents, Rasmussen (1997, p. 1) poses the question of whether “we actually have adequate models of accident causation in the present dynamic society”. His highly influential research studies describe the challenge faced by safety investigators when dealing with complex socio-technical systems. Leveson (2016) observes that Rasmussen’s fresh approach resulted in a paradigm change in accident research. After the 1979 Three Mile Island nuclear accident, the safety community was confronted by evidence that complex systems can fail in ways not foreseen by the designer and regulator (Rogovin, 1980). That realization led to Perrow’s (1999) seminal work on normal accident theory. Interestingly, in the immediate aftermath of the accident there was near consensus that it was caused by operator error and the main research focus remained on fixing the mental model of humans controlling complex systems by generating the right feedback and symbols (Cook, 2012; Rasmussen, 1987).

In airline accident theory, the 1989 Air Ontario F28 crash at Dryden was a similar watershed moment (Moshansky, 1992). Without Cockpit Voice Recorder (CVR) or Flight Data Recorder (FDR) evidence, the original investigation concluded that the cause of the aircraft attempting a failed take-off with heavy ice buildup on the wings was pilot error (Reason, 2008). Helmreich et al. (1999) looked at the many systemic and human factors involved in the crash and invented CRM as a solution. The Moshansky (1992) report came to an entirely different conclusion: Dryden was caused by the Canadian airline system, including the operator, the regulator, government policies, and other systemic factors.

Evaluating NASA’s disastrous Space Shuttle Challenger launch decision, Vaughan (1996) builds on the sociological approach and describes her theory on how complex bureaucratic machines slowly, but inevitably, drift into failure. She considers the Challenger disaster as an example of a normal accident. Countering her claim, Perrow (2004) stated that Challenger was a typical component failure accident where management was the key component that failed, while Weick (1998) supported her opinion about how the crisis unfolded. Vaughan's (1996) drift theory remains influential to this day, especially in the field of organisational risk management and resilience studies (Dekker, 2005; Woods, 2003).

The fundamental difference between the above accident theories is how failure is seen, evaluated, and attributed. In an aviation context, before the jet age, a Cartesian (mechanical) view of the world was a reasonably close approximation to describe the aircraft and its critical systems (Cook, 1998). Leveson (2011, 2012) highlights that the Western way of decomposing the system into individual components resulted in describing even the pilot and maintenance engineers as information processing units that can and do fail. It may have been a perfect analogy to support the designer’s goals and means analysis, but that model simply does not work for complex systems. Attributing accidents to a root cause is wrong and misleading for today’s highly automated systems that operate in hazardous settings (Leveson, 2011, 2012; Rasmussen, 1997).

Problems With Causal Reasoning. Rasmussen and Svedung (2000) observe that socio-technical systems are increasingly complex and to be able to analyse the behaviour of the system's building blocks, connections, events, and decisions, the phenomenon needs to be separated from that complex world. Humans experience the world in a continuous flow and the accident analyst needs to create both structural and functional models of that world before looking for causal explanations. Dismukes et al. (2007) add that the abstraction is only meaningful if it is done at a level and in a context that is familiar to the investigator. Explanations are not true or false, only more or less *plausible* in each context. In modern system safety terms, explanations are given in probabilistic terms, described as a chance combination of errors and multiple situational factors.

Causal explanations are also constrained by the observer's frame of reference and the ambiguity of stopping rules applied during causal analysis. It is tempting to focus on things that go wrong and components that fail, but those are only symptoms apparent on the surface (Dekker & Woods, 2002). Normal (standard) operations should not be taken for granted by the analyst. Conditions change, which may have an unexpected or cascading effect on the system (Rasmussen & Svedung, 2000). By necessity, accident causal trees only record a single case of that propagation, which may or may not have happened that way during the occurrence. That is not the same as describing a relational structure in the accident model (Dismukes et al., 2007; Leveson, 2011).

Another important caveat to causal theory is the problem of causal asymmetry. When necessary and sufficient conditions are not met, the causal effect may still occur (Lewis, 2008, Pearl, 2000). Counterfactual techniques represent one of the few successful methods for reducing bias in confidence judgements (Sloman, 2008). That is why formal and semi-formal investigation methods try to reverse engineer the causal relationships embedded in investigation reports by counterfactual argumentation (Burns, 2000; Johnson, 2003; Ladkin, 2000). Unfortunately, the problem of causal asymmetry is a major challenge for formal methods that are explored in the following paragraphs.

Formal Methods. Natural language arguments form the basis of accident reports which conform to the ICAO (2020b) recommended standard. Natural language descriptions offer many benefits when describing causal reasoning and findings, including a rich context, assumptions, probabilities on a wider range of scale, and finer details (Ladkin & Loer, 1999). At the same time, if the reader must read between the lines, looking for implicit assumptions, it can be difficult to understand why the investigation reached a particular conclusion (Johnson, 1997a, 2003). Over the years, many attempts were made to supplement ICAO standard accident reports with formal analysis methods (Burns, 2000; Johnson & Holloway, 2003; Ladkin, 2000, Pearl, 2000). The core idea was to introduce logical reasoning for better structuring critical information and arguments in the report, looking for missing information, and supporting a more coherent summary of investigation findings. Most formal methods rely on applying counterfactual logic for independently checking the integrity of causal reasoning embedded in accident reports (Lewis, 2008; Pearl, 2000).

While some formal methods – such as Conclusion Analysis and Evidence diagrams (Johnson, 1997b; Johnson & Holloway, 2003), Multilinear Event Sequencing (Benner, 1975), or Why-Because Analysis (Ladkin, 2000) – have been successfully applied as an objective a posteriori reasoning method and identified significant causal reasoning mistakes in controversial accident report findings, these methods are complex and time-consuming. The main limitation remains the fact that the analysis is dependent on the original accident investigation, as well as the initial set of causal factors identified by the investigators (Burns, 2000). The conventional process, however, relies on witness interviews and factual evidence gathering, sources which are highly susceptible to subjectivity and filtering (ICAO, 2015). Formal methods typically do not structure facts into causal categories, and as such, their value in identifying safety recommendations is limited (Burns, 2000; Leveson, 2001).

Holloway and Johnson (2006) posit that a counterfactual definition is too narrow and too broad at the same time. Too narrow, as it eliminates organisational factors from the causal picture. And too broad, as it is difficult to make a distinction between causes and conditions. Natural language allows for exceptions in describing causal relations, which are very difficult to codify in a deterministic form (Pearl, 2000). In summary, due to the significant limitations, investigators should not rely on formal proof for causal reasoning.

2.7.3 Systems Approach

Leveson (2012, 2016) finds that while modern systems are increasingly complex, traditional models of causality fail to explain what causes catastrophic accidents in large-scale industrial applications. Some accidents were not supposed to happen. Cook (2012) and Leveson (2012) both observe that the need for a systems approach to safety was identified in military programs, biology, the nuclear industry, and commercial aviation at roughly the same time.

The final section of the literature review outlines the origins, key concepts, and practical application of accident models that are based on system theory principles. Apart from normal accident and high reliability theories, the models described here all originate from the CSE school initiated by Rasmussen, Hollnagel, and Woods (Le Coze, 2022). From the CSE school, the section focuses on the key concepts of resilience, functional resonance, control theory, and complex adaptive systems.

Seminal contributions from Rasmussen and Dekker are not repeated in this section.

Normal Accidents. As mentioned earlier, Perrow (1981) introduced the concept of normal accidents when describing the 1979 Three Mile Island nuclear accident. He argues that no matter how hard we might try, any tightly coupled system with interactive complexity is bound to fail eventually, where interactive complexity refers to a system with many nonlinear interactions. Perrow's (1999) analysis concludes that potentially catastrophic systems must either be shut down, or society must accept the risk of major safety losses, including loss of life at a major scale, when normal accidents occur.

Perrow's (1981) view did more than try to characterize systems with a catastrophic potential. The new accident theory was the starting point for more research into how large disasters unfold and to understand less complex accidents where key organisational components failed. Typical examples cited in the literature are disasters in nuclear and chemical processing plants (Leveson, 1995), military operations (Sagan, 2004; Snook, 2000), space flight (Vaughan, 1996), or air traffic system meltdowns (Dekker & Pitzer, 2016; Weick, 1990).

High Reliability Organizations. Around the same time, another major school of thought suggested that High Reliability Organizations (HRO), where all components of the system operate reliably, can achieve, and maintain safety. The underlying theory evolves around the reliability and culture of complex organizations and their technologies. The primary traits of such an HRO are redundancy, safety as the primary objective, a culture of reliability, continuous operations, and a strong capability to learn (La Porte, 1996; Roberts, 1990; Weick, 1987).

Leveson et al. (2009) identify a fundamental flaw in HRO theory, highlighting that safety and reliability are different system properties. One does not imply or require the other, as a system can be *reliable and unsafe* or *safe and unreliable*. There are many examples of complex systems where accidents happen due to the interaction of perfectly functioning and reliable components, or large organisations that are unable to prevent or learn from mistakes (Leveson, 1995; Leveson et al., 2009; Sagan, 1993).

Resilience Engineering. Hollnagel (2011) introduces the concept of resilience engineering, which defines safety as a socio-technical system's ability to succeed under varying conditions. The concept intends to replace the traditional search for accident causes with an understanding of how the system's performance failed. Resilient organisations were meant to be built on four abilities: the ability to respond to events, monitor developments, learn from past failures, and anticipate future threats. Until recently, the concept of resilience remained elusive with few practical resilience engineering examples published in the literature (Cook & Long, 2021).

Woods and Cook (2017) explain that by defining resilience as an understanding of how well the system adapts and to what range or source of variation it adapts, we can determine what constitutes an undesirable reduction in adaptive capability. Decompensation is an important concept in that adaptive response cycle. Automation can compensate for an increased disturbance, and if successful, it can partially mask the disturbance. A decompensation event happens when automation cannot compensate any more. The key question is whether the pilots (as supervisors) detect the developing problem during the initial phase. If they miss the signs from automation (the base controller) that the machine is approaching capability limits, the system can be pushed over its safe boundaries (Woods & Branlat, 2017).

Control Theory. Leveson (2014) approaches system accidents from the angle of supervisory control theory. Her observations about software-intensive systems are directly relevant to the non-routine safety problem explored in this study. Hazard analysis methods were not updated to match the new digital work environment described by Mosier (2008) and the new types of accident scenarios introduced. During design and certification, traditional techniques miss many possible scenarios, especially those related to software or humans (Leveson, 2014).

In response, Leveson (2004, 2012) developed the System–Theoretic Accident Model and Processes (STAMP), a new accident model that aims to provide a better fit for modern designs by treating safety as a control problem, not as a failure problem. Causal Analysis based on STAMP (CAST) is an a posteriori accident analysis technique, which looks for inadequate control loops. STAMP (CAST) has been successfully applied to several industrial and aerospace accidents, including high-profile airline investigations such as the 2014 Asiana B777 accident at San Francisco (Thomas & Malmquist, 2016), the 2009 Air France A330 crash over the Atlantic Ocean (Malmquist & Leveson, 2019), and the 2006 Comair CRJ-100 accident at Lexington (Nelson, 2008b).

Functional Resonance. Hollnagel and Goteman (2004) describe the Functional Resonance Analysis Model (FRAM). FRAM relies on the concept of stochastic resonance when describing the dynamic nature of accidents. The goal of a FRAM analysis is to monitor and control performance variability in essential system functions. According to Hollnagel (2016), FRAM achieves that goal by identifying barriers for variability (also referred to as damping factors) and specifying the required level of performance monitoring. There were repeated attempts at applying FRAM in a retrospective analysis of routine airline accidents and incidents, such as the 2000 Alaska Airlines MD-83 accident in California and the 1997 AOM Minerve MD-83 incident at Paris (Nouvel, Travadel, & Hollnagel, 2007; Woltjer & Hollnagel, 2007). The case studies struggled to highlight the key events, the role of key actors, and the dependencies involved.

Leveson (2020) is very critical of Hollnagel's (2016) approach to performance variability. She highlights that a safe system design must recognise the inherent conflict between production and safety goals and eliminate or minimize those conflicts. When operational performance (the aircraft, the pilot, software, management, etc.) varies outside safe boundaries, the system design must still maintain adequate safety levels. FRAM uses hexagons to identify system functions, with each hexagon described by time, input, output, control, preconditions, and resources required for executing that function (Hollnagel, 2016). Leveson (2020) is correct to highlight that in system engineering terms, that description is equivalent to a system specification, not an accident model.

Theory of Graceful Extensibility. Woods (2018) offers an extension of resilience engineering principles through his Theory of Graceful Extensibility. The formal theory is based on earlier notions of complex adaptive systems and cognitive patterns (Woods, 2015; Woods & Hollnagel, 2006). The key concept of Woods' (2018) new theory is *graceful extensibility*, defined as the system's ability to adapt when surprise events push the system up to its boundaries. The theory predicts that resilience will appear when individual adaptive units exhaust their own adaptive capacity and need to obtain additional capacity from other adaptive units in their network. Cook and Long (2021) provide an empirical case study for technical incidents in internet-facing critical business operations (airline systems, hospitals, stock exchanges, etc) and confirm Woods' (2018) prediction that sharing adaptive capacity enables the system to respond to severe challenges in a graceful manner.

In a related contribution, Woods and Branlat (2017) describe three basic patterns in how adaptive systems fail. The basic patterns are decompensation, adaptive units working at cross-purposes, and outdated behaviours. Decompensation refers to the effect of reaching saturation when the system cannot adapt to disturbances. Working at cross-purposes refers to units working toward local optimums that are globally maladaptive. Finally, outdated behaviours are exhibited by pursuing outdated models and plans.

2.7.4 Section Summary

The final section of the literature review explored the theoretical foundations of air safety investigations, especially the suitability of underlying accident models for identifying causal factors involved in complex occurrences. While the stated objective of ICAO (2011, 2020b) recommended investigation standards is to identify systemic causes, including organisational and human factors involved in the accident, the framework is built on traditional RCA methods. A range of formal and semi-formal methods were put forward by the research community to supplement natural language arguments in ICAO (2020b) standard investigation reports, but significant limitations prevent their practical application (Burns, 2000; Holloway & Johnson, 2006; Leveson, 2001).

Approaching airline accidents from a systems view appears to be more appropriate but poses the challenge of applying more complex accident models. System-theoretical causal analysis has been successfully applied as an a posteriori method, but the perceived complexity and steep learning curve associated with the new model have prevented widespread uptake by the air safety investigation community so far (Underwood & Waterson, 2013). While the concept of resilience and adaptive capability is relatively new, the idea of looking for decompensation patterns when assessing how a highly automated aircraft system responds to disruptions and challenges offers a promising theoretical framework. Woods' (2018) extension to resilience engineering theory and Woods and Branlat's (2017) definition of basic patterns in how adaptive systems fail is supported by a successful case study into significant technical failures in critical settings (Cook & Long, 2021). This new framework is suitable for learning more about patterns involved in check flight accidents and incidents.

2.8 Summary

The literature review identified the following gaps in the context of post-maintenance check flight operations:

The scoping review failed to uncover any safety statistics or readily available safety data for non-routine flight occurrences. Most commercial aviation safety reports exclude occurrences during non-routine flying operations. A notable exception is the annual statistical summary published by Boeing (2022), which provides a simple tally of hull losses and fatal accidents for maintenance test, ferry, positioning, training, and demonstration flights. It follows that the project first must generate research data by building a catalogue of non-routine accidents and incidents.

Other than Poprawa’s (2015) rough estimates, the literature base provides no guidance in terms of the actual check flight safety risk levels. Due to the lack of historical data and research on the topic, it is not clear whether the airline industry is fully aware of actual system safety risk it is exposed to during non-routine flight operations. Publications released by safety boards, regulators, airframe manufacturers, and the very few independent articles published on the topic, all reflect a narrow view that check flight safety is solely an operational airworthiness problem and should be treated as such. This approach contradicts the evidence obtained through studies of human-machine interactions in highly automated settings.

Traditional accident models and causal analysis frameworks are not suitable for describing the systemic causal factors involved in complex systems, such as highly automated airliners. Causal findings in root cause investigations are arbitrary. Due to inherent limitations, formal and semi-formal methods cannot be applied to complex systems. Accident models based on systems engineering principles are readily available, but some approaches are resource-prohibitive. Woods' (2015, 2018) extension to resilience engineering theory is suitable for uncovering the basic adaptive patterns involved in check flight accidents and incidents.

The purpose of this research study is to develop causal explanations for maintenance test flight occurrences. The scoping review identified a few examples of air safety investigation reports in grey literature that can provide a starting point for addressing this objective. Gephart's (1993) important contribution to organisational sensemaking demonstrated that a textual approach to safety investigations can deliver significant insights into public inquiries, causal explanations, and the way blame is apportioned during the process. This realization led to posing the first research question:

Research Question 1

What causes and contributing factors are annotated in relevant check flight investigation reports?

Due to the critical gaps mentioned earlier, the first question cannot be tested without building up the necessary research data. For that purpose, the study elected to employ a systematic review of government archives (grey literature) as the first step in the methodology, followed by Holloway and Johnson's (2004) approach to data analysis, as explained in the next chapter.

A second purpose of the study is to explore the relationship between the safety response implemented by the industry and the common patterns and cross-cutting themes identified across multiple critical incidents. The safety response is known from primary and secondary sources identified during the scoping review (Airbus, 2015; Regulation (EU) 2019/1384, 2019; FAA 2008b, 2016; FSF, 2011a, 2011b; Poprawa, 2015). Common causal patterns and themes are not known, and as such, the objective can be achieved by adding two more research questions to the study:

Research Question 2

Are there any common patterns in check flight occurrences? If there are common patterns, what do they reveal about the deeper structure of check flight accidents?

Research Question 3

Is there an alignment between underlying patterns and the risk control framework implemented by the airline industry?

The second and third research questions adopt Woods' (2015, 2018) formal extension of resilience engineering, supported by a robust accident analysis framework developed by Dismukes et al. (2007) for revisiting air safety investigation reports. The next chapter provides a detailed overview of the qualitative research design and the methodology applied when testing the research questions.

CHAPTER 3 – METHODOLOGY

3.1 Introduction

The primary goal of this study is to answer the research questions that relate to maintenance test flight accident causation. Separate methods were utilized to this end. This chapter presents the overall qualitative research framework and the methodology as employed to test the research questions. Accordingly, the methodology chapter is organised into the following sections: a) research design, b) data sources and collection c) initial data analysis, d) embedded multi-case study, and e) validity threats.

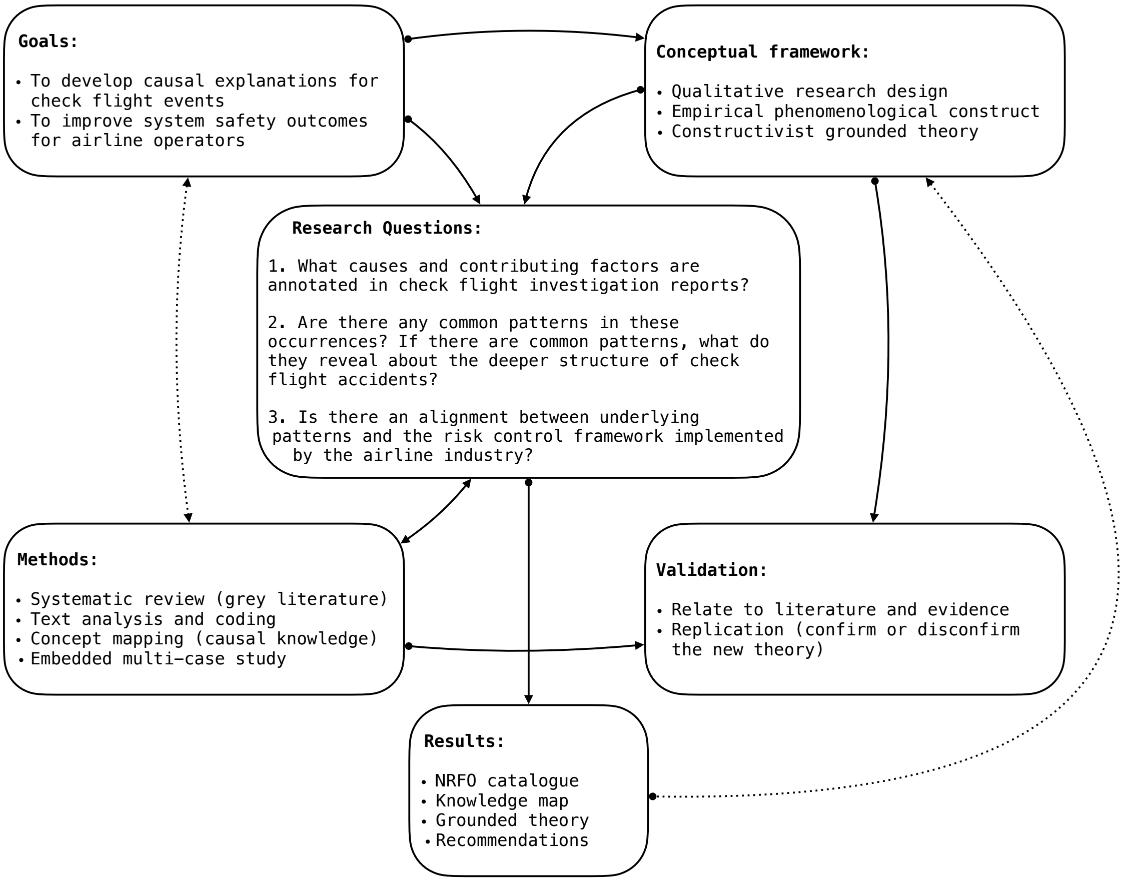
3.2 Research Design

The qualitative research framework applied in this study follows Maxwell's (2009) interactive model that has five main components. Figure 5 illustrates the actual structure of the overall research project and the relationships between the main components. This section provides an overview of each component that are essential to form an integrated and coherent structure, including the goals, the conceptual framework, the research questions, the method, and validation challenges.

The *main goals* of the study are to develop causal explanations for check flight events and to improve system safety outcomes during non-routine airline flight operations. Admittedly, there is a duality in these practical goals. On the one hand, the study aims to move beyond outdated concepts of causality, identifying unanticipated phenomena and developing a new grounded theory of accident causation. On the other hand, the goal is to generate a result that is experientially credible, both to airline operators and airframe designers. As such, it was important to select methods that achieve sufficient level of detail and credibility in support of the system safety recommendations put forward (Maxwell, 2009).

Figure 5

Research Design



The *conceptual framework* component refers to the author's tentative theory, generated through studying past accidents, scoping the literature, and experiential knowledge (Strauss, 1987; Strauss & Corbin, 1990). The empirical phenomenological part refers to first-order constructs identified in accident investigation reports as well as the second-order constructs created for simplified causal maps (Aspers, 2009). The key organizing element of the framework is the idea that a new constructivist grounded theory can be formulated from an embedded multi-case study of maintenance check flight incidents. The new theory is grounded in case stories and informed by tentative constructs identified in existing literature (Eisenhardt, 1989; Maxwell, 2009).

The three *research questions* are at the heart of the framework, directly connecting all other components. In line with the inductive nature of qualitative studies, the research questions were only drafted once primary goals and the conceptual framework had been robust enough. The relatively small sample size and varying quality of historical investigation records limited what was feasible in terms of data analysis (methods selected). In turn, this limitation informed the final version of research questions.

Methods have been preselected by pragmatic considerations. Prior to this study, there were only fragmented data elements for non-routine flight occurrences. A systematic review of historical records served as a reliable and repeatable method for assembling data for this research. The initial text analysis and coding process was borrowed from a comparable NTSB study of commercial airline accidents (Holloway & Johnson, 2004). Concept mapping is one of the few techniques suitable for communicating the complex first- and second-order constructs and causal relationships involved in this study (Hoffman & Lintern, 2006). The initial data analysis part provided the response to research question one. Finally, Eisenhardt's (1989) multi-case study method was selected to generate a novel theory for check flight accident causation, testing research questions two and three.

Yin (2014) identifies *reliability*, *construct validity*, and *external and internal validity* as major challenges for multiple case studies. As explained in the last section of this chapter, the study employed two main strategies in addressing validity threats. The multiple case design enabled Yin's (2014) replication logic that served to confirm or disconfirm inferences drawn for previous cases. In addition, the case studies constantly iterated between the rich context provided by research notes and whatever similar findings were available in the literature or through empirical evidence (Eisenhardt, 1989, 2021).

The *results* component is provided for the purposes of clarity. The initial data analysis part led to the creation of the *NRFO catalogue* (research data) and a *knowledge map* that summed up the response to question one within the bounds of historical investigations. The cross-case analysis identified *common patterns* in MCF events that led to a *new grounded theory*. Finally, the study adopted *practical recommendations* by comparing the existing plan and prepare approach with the results of this research project. The rest of this chapter revisits the applied methods and techniques in more detail.

3.3 Data Sources and Collection

3.3.1 Systematic Review

The scoping review identified that a comprehensive list of non-routine flight investigations is not available in published literature. Historical records in grey literature, including records maintained by government safety agencies and independent aviation safety databases, are a valid and reliable data source for this purpose. Therefore, the research study embedded a systematic review as a necessary precursor before the data analysis phase. A systematic review served as a reliable and repeatable method for sourcing air safety investigation reports issued worldwide.

3.3.2 Source Data for the Systematic Review

The primary data source for accident and incident investigation reports is grey literature through government transport safety agencies.

The following six safety bureau websites were selected, in alphabetical order:

- Air Accidents Investigation Branch (United Kingdom) (AAIB, n.d.)
- Australian Transport Safety Bureau (ATSB, n.d.)
- Bundesstelle fur Flugunfalluntersuchung (BFU), the German Federal Bureau of Aircraft Accident Investigation in Germany (BFU, n.d.)
- Bureau d'Enquetes et d'Analyses, the Bureau of Enquiry and Analysis for Civil Aviation Safety in France (BEA, n.d.)
- National Transportation Safety Board in the United States (NTSB, n.d.) and
- Transport Accident Investigation Commission (TAIC) in New Zealand (TAIC, n.d.)

Independent aviation safety databases provide a secondary source of research data in grey literature. Additional records were sourced from the following public databases available on the internet:

- Aviation Safety Network (ASN) data maintained by the FSF (FSF, n.d.)
- Aviation Herald (AvH) database (AvH, n.d.)

- Flight Safety Systems (FSS) LLC database (FSS, n.d.) and
- The Bureau of Aircraft Accidents Archives (B3A, n.d.)

Internet records were only included in the catalogue, if the safety event was traceable to an ICAO member state's investigation report, safety alert, or occurrence database entry.

3.3.3 Eligibility for Inclusion

Fourth generation commercial jets entered revenue service in 1988 with the introduction of the Airbus A320 family. Therefore, a 1 January 1988 starting date was selected when assembling the research dataset, ensuring that a full in-service history is captured for the latest generation.

As outlined in chapter one, only Western-built fixed-wing commercial jets were included in the systematic review. Delimiting the research scope to this population increases the reliability of the study. There are no reliable data sources for Eastern-built models and the only verifiable non-routine safety statistics are limited to commercial airliners (Boeing, 2022). As such, investigation records were filtered by make and model, applying a Maximum Take-off Weight (MTOW) over 60,000 pounds (27,200 kg) rule, which broadly represents commercial airliners in mainline passenger and freight transport operations.

3.3.4 Exclusion Criteria

In addition to Eastern-built fixed-wing airplanes and any Western-built fixed-wing airplane with an MTOW less than, or equal to, 60,000 lbs, the study also excluded Western-built turboprop powered fixed-wing airplanes; helicopters, gyrocopters, airships, balloons, gliders and hang gliders.

3.3.5 Search Strategy

The primary search iteration was completed by extracting accident and incident reports from the selected major safety boards' online databases. The secondary search iteration was achieved by interrogating the non-government aviation safety databases, then manually matching the records with those found in the first iteration. Any duplicate records were removed from the dataset. The remaining non-routine events were traced to the relevant ICAO Contracting State's investigation report or safety alert. The relevant transport safety databases, website search criteria, and additional filters utilized in the systematic review are available in Appendix A.

3.3.6 MCF Events

As highlighted in the introduction, non-routine flying is a broad operational category. Therefore, the NRFO catalogue includes safety occurrences during positioning and ferry flights, training flights, demonstration, end-of-lease, and maintenance check flights. As a final step of the systematic review, MCF events must be extracted from the NRFO library by exclusion. First, all declared positioning, ferry, and training flights were removed from the list in Appendix B. The remaining records were reviewed line-by-line to identify and remove demonstration flights. Finally, all non-routine events were classified as an *operational or functional check flight* event to support the data analysis phase described below.

4.3 Initial Data Analysis

This section introduces the initial data analysis steps, including the way causal constructs were built up from investigation reports, followed by causal categories that were identified when short-form text labels were allocated to sufficiently similar causal constructs, and finally, how the underlying causal logic was translated to a graphical representation.

3.4.1 Causal Constructs

While the preferred safety investigation methods may be different, investigation reports produced by transport safety agencies conform to ICAO (2020a) Annex 13 standards. A common feature of these reports is a heavy reliance on natural text when describing the events leading up to the safety loss, including the investigation team’s analysis, findings, annotated causes and contributing factors. Providing a description in the investigator’s primary language, especially when conveyed to a person who understands that same language, is a powerful tool in communicating the proposed causal structure embedded in the report. Some of that context, however, is lost when the report needs to be translated to English, the common language of the international aviation safety community.

More importantly, ICAO (2020b) standards do not prescribe a common terminology for annotating probable causes and contributing factors in investigation reports. The ICAO (2020b) investigation manual offers examples of commonly used findings; however, it is up to the investigator-in-charge to ensure that findings are pertinent and valid in each context (p. 52). As a result, dissimilar labels can be assigned to seemingly identical causal linkages in safety investigation reports, or the same label may refer to a cause in one report and a causal factor in another investigation. Therefore, there is only limited value in directly comparing causal labels annotated in MCF investigation reports.

Holloway and Johnson (2004) developed a data analysis method for the purposes of comparing causes annotated in NTSB airline accident investigation reports. This study adopted the same phenomenological analysis that involves building up the textual and structural context from horizontal clusters, which are then delimited to invariant themes (Patton, 2002). The synthesis of those initial results provided the framework for portraying overall temporal and structural characteristics of check flight investigation reports in a knowledge map, as explained later in this chapter. In the first instance, causal constructs were extracted from MCF investigation reports by a simple text search. Most transport safety agencies annotate (probable) causes and contributing factors in their investigation summary. Where the safety agency's policy is not to nominate a cause, systemic causal factors are extracted from the report. This initial analysis step reveals first-order logical constructs that reflect the investigator's theory about causal relationships (Aspers, 2009).

3.4.2 Causal Categories

Next, the original long-form text was replaced by short-form causal labels. The labels were limited to a few keywords which would be sufficient to describe the investigator's causal findings for other safety investigators. Substantially similar causes and factors were assigned a causal category based on these short-form labels. Holloway and Johnson's (2004) method applies the same steps in their causal analysis of NTSB investigation reports.

In the context of causal reasoning, natural language also offers the ability to express the whole range of potential causal relationships, from almost certain, through neutral, to negative causal effects. This characteristic was essential in building and evaluating cognitive maps from accident investigation reports, as described in the next section.

3.4.3 Causal Logic

Concept maps are directed graphs, where nodes (concepts) are linked by edges (arrows). When drawing these maps, the focus is on describing the relationship between the nodes, supported by labels (typically action verbs) assigned to the directed edges. Unlike in a mind map, a concept node can have multiple parent nodes, a flexible hierarchy which allows for illustrating interconnected graphs (Hoffman & Lintern, 2006). Concept mapping originates from education studies, supported by decades of research and development. The technique has widespread application in eliciting and representing expert domain knowledge (Hoffman & Lintern, 2006). Concept maps are ideal for describing connections between qualitative research elements for the purposes of theory development. Concept maps turned out to be an efficient visual communication aid for clarifying existing theories or developing new ones (Maxwell, 2013; Miles & Huberman, 1994, Wiggins and Glass, 2013).

Axelrod (2015) introduced *cognitive maps* for representing social scientific knowledge. Cognitive maps are signed digraphs where nodes are variable concepts and edges indicate *causal increase or decrease* between nodes. Cognitive maps facilitate documentary coding of historical records, such as investigation reports. As outlined in the previous section, there is an inherent level of uncertainty in describing causal relations in investigation reports.

Kosko (1986) introduced Fuzzy Cognitive Maps (FCMs) for representing uncertain relations between causal concepts. In an FCM, causal descriptions are supported by matrix representation (connection matrix) and matrix operations. Any number of FCMs can be combined by simply adding the corresponding connection matrices. This property allows growing knowledge bases by simple adaptive inference.

Simple FCM graphs are suitable for representing check flight investigation results. A simple FCM has causal edge weights in a {-1, 0, +1} trivalent set. All causality is nonfuzzy. Causal descriptions in ICAO (2020b) standard investigation reports can be better approximated by a {0, +0.5, +1} set, where a +1 value annotates a *(probable) cause*, while a +0.5 value signs a *contributing factor*. Neutral causal links are self-explanatory, indicated by the lack of arrows between concept nodes (Kosko, 1986). For the purposes of clarity, the study refers to simple FCM maps that directly translate ICAO standard investigation reports as *trivalent causal maps*.

A trivalent causal map was drawn up for each check flight event. The maps were drawn at the level of causal categories identified in the previous step. Some finer details about systemic causes, contributing factors, and the interactions between those elements remain hidden at this higher abstraction level. As such, trivalent causal maps are not eqivalent to a complete accident model. The individual graphs may be more appropriately described as sharing the analyst's mental model, i.e. the analyst trying to make sense of investigation outcomes (Holloway & Johnson, 2004; Weick, 1979).

3.4.4 Knowledge Map

The final step of the initial data analysis phase exploited a key benefit offered by FCM representation. An arbitrary number of FCM graphs can be naturally combined into a non-simple graph that represents overall causal magnitudes. FCMs are combined by first creating augmented connection matrices. That is a simple mathematical transformation trick by making sure that each individual connection matrix has the same number of rows and columns. If the total number of concepts is n, then the connection matrix is augmented to an n x n square matrix (Kosko, 1986, 1988).

The combined FCM matrix F is found by adding up the augmented connection matrices of each trivalent causal map:

$$F = F_1 + F_2 + \ldots + F_k$$

where k is the total number of trivalent causal maps in the sample. The result is a non-simple (fuzzy) graph. Edges in the combined graph are normalized by 1/k to account for the number of data sources (Kosko, 1988).

For the purposes of clarity, the study refers to the combined non-simple (fuzzy) graph as the *knowledge map,* to differentiate between individual contributions from causal maps and the overall relationships and causal magnitudes revealed in the collective representation. In Asper's (2009) empirical terms, the knowledge map is a graph representation of the author's second-order logical construction that is grounded in the first-order constructs elicited from MCF investigation reports. The knowledge map is a valuable contribution through a unique representation where accident investigators' perspective is central to the analysis (Aspers, 2009). Leveson (2012) finds that underlying accident models influence what causal factors are considered during a safety analysis. It follows that the knowledge map is bound by the investigation methods applied, and as such, the map inherits the deficiencies and logical inconsistencies embedded in constituent reports.

Most concerning of those shortfalls is a typical stopping rule applied during root cause or event chain type investigations. Rasmussen (1990) suggests that nominating human error as a root cause is directly related to the difficulty involved in backtracking event chains through human operators. Whenever operator error was annotated as a root cause, rather than treated as a symptom of underlying systemic issues, the trivalent causal map inherits the cognitive biases involved in blaming the flight crew for the safety loss. The same biases are carried forward to the knowledge map.

Learning from past MCF investigations is very important, but the safety response should not be limited to fixing individual causal trees drawn up for arbitrarily selected root causes. It is extremely unlikely that the same accident or incident pattern would unfold in the same manner, especially when complex systemic causes and factors are involved. Johnson and Holloway (2003) refer to this problem as a counterfactual paradox that justifies the need to look beyond classical logic when mishaps are analysed.

3.5 Extended Data Analysis

3.5.1 Embedded Multiple Case Design

As highlighted in previous paragraphs, the knowledge map can only provide a limited view on causal reasoning. A comparative case study is a powerful qualitative tool for generating causal explanations that can either be employed as a primary framework, or embedded in a broader research design (Goodrick, 2019). This study elected to adopt the latter approach, an *embedded multiple case* design that relies on several units of analysis. An embedded multiple case design is relatively complex, but it provides a rich context for explaining systemic behaviour (Yin, 2014).

Yin (2014, p. 16) defines case study as an empirical inquiry which investigates a phenomenon in its real-life context. Flyvbjerg (2011, p. 301) prefers the Merriam-Webster definition of "An intensive analysis of an individual unit (as a person or community) stressing developmental factors in relation to environment.", stressing that the decisive factor in defining case studies is the unit of analysis and the setting of its boundaries. Mumford et al. (2018) highlight that a case reflects a series of events in a real-world context, where events refer to Pillemer's (1998) definition as "an action, or set of actions, taken with respect to environmental contingencies".

There are several approaches to case studies. Grounded theory is a systematic method for constructing theories through analysing qualitative data and proposes no prior literature review (Glaser & Strauss, 1967). Yin (2014) proposes a more directed study with a priori constructs. Eisenhardt (1989, 2021) describes an inductive case-oriented process that is especially suited for generating novel theories. Charmaz (2017, 2017a) offers constructivist grounded theory, a reflexive approach to critical qualitative inquiry. This study elected to adopt a multiple case *Eisenhardt Method* for the purposes of testing research questions two and three. Eisenhardt's (1989) landmark contribution created a highly iterative process that delivers testable, and empirically valid, grounded theories. It relies on Yin's (2014) replication logic and Glaser & Strauss'

(1967) constant comparison of theory and data. The following sections provide a roadmap to the method applied in this study.

3.5.2 Eisenhardt Method

The roadmap presented here synthesizes the actual method from research questions to reaching closure, and the theories that influenced the data collection and analysis phases, shaping and extending the new grounded theory for check flight accident causation. The roadmap follows the process steps outlined in Eisenhardt's (1989, 2021) seminal contributions.

Research Focus and Piror Constructs. The study aims to uncover the dynamic causal relationships involved in maintenance test flight accidents. The overarching research objective is to discover under what conditions, and through what causal paths, maintenance test flight accidents unfold. As recommended by Eisenhardt (1989), the case study is informed by tentative constructs that were identified through the causal mapping process. At this stage of the analysis, categories along stronger causal patterns in the knowledge map are important, as they can potentially emerge as relevant to building the new causal theory.

Selecting Cases. Case selection followed a theoretical sampling plan from a small population of 21 MCF events. The goal of theoretical sampling is to choose cases that are likely to replicate or extend theory by mapping conceptual categories (Eiasenhardt, 1989). The sample was not random, as the selection reflected the objective to extend the theory to all airliner generations. The sample was designed to include polar types, in other words failed and successful recoveries that ended in accident(s) or incident(s).

Entering the Field. Data collection entirely relied on historical investigation reports sourced from government archives during the systematic review process. In contrast to Gephart's (1993) ethnomethodological approach, it was not feasible to conduct interviews with persons involved in the incident or the safety inquiry. Of special note is the author’s extensive maintenance test flight experience that helped with corroborating findings from qualitative evidence.

Context rich research notes served as the most important means of iterating between data collection and analysis. The research notes form the initial part of the case study chapters and follow the same explicit theoretical structure. Observations include any noteworthy events leading up to the incident flight, a nominated transition point, and significant events pre- and post-transition. Transition refers to a critical decision point where the check flight sequence starts shifting from the original test plan to a recovery phase.

In lieu of the investigation report’s focus on probable causes, the pre- and post-transition narrative is guided by two focus questions:

Focus Question 1

Why did it make sense for the crew to continue the test program?

Focus Question 2

Why did the recovery attempt work / fail to work?

The primary benefit of structuring the case study around the new focus questions is the credit given to both positive and negative causal links, where a negative link refers to a causal decrease, not a negative safety impact. In this research study, causal decrease is more appropriately described as not supporting the sensemaking or decision making process (Klein, 2017; Klein & Klinger, 1991; Klein et al., 1993; Klein & Wright, 2016). By introducing the concept of causal decrease, case study notes require an extended version of causal maps. Pre- and post-transition causal maps utilize a pentavalent logic that consists of a {-1; -0.5; 0; +0.5; +1} value set. The new {-1} label reflects a strong negative causal weight, while a {-0.5} label reflects a degree of uncertainty (haziness) when a causal decrease is assigned to a relationship between relevant nodes.

Within-Case Analysis. Eisenhardt (1989, 2021) does not prescribe a standard format for within-case analysis, but rather only highlights the need to reveal the unique patterns of each case before patterns are generalised across cases. In addition to the chronological sequence of key events, this study also used observations about adaptive failures to organise case data.

Within-case analysis builds on Woods' (2018) modified theory and sub-theorems about resilient systems. The focus is on three basic patterns in how adaptive systems fail: a) decompensation, b) locally adaptive and globally maladaptive responses, and c) outdated plans or outdated behaviours (Woods & Branlat, 2017). To achieve that aim without building a complex network model, the study assumes that a simple interaction model can describe the main adaptive units in each case.

Analysis data is provided in a tabular format, followed by a narrative description of observed patterns. The study introduces time as a critical dimension for failure analysis. Following the modified resilience terms introduced by Woods (2018), the analysis follows key challenges and surprises (events), the actions deployed by controllers (response), the risk of saturation (compensation), any potential misalignment or mismatch between the units (coordination), and any evidence of being stuck in outdated models when the environment has changed, or a new disturbance was introduced (revised model).

Tentative Concepts and Categories. The selection of tentative concepts and categories was informed by the research questions and existing theory elements in the literature. The research objective defines the outcome variable as the *magnitude of safety loss*, reflected in an accident (fatal), accident (hull loss), or incident outcome. The safety loss variable was later extended by adding *recovery (full, partial, or failed)* as an additional outcome measure to the comparison table.

Existing theory introduced tentative concepts, such as, *crew training*, *communication between pilots and maintenance engineers,* and *system or component failures* (Poprawa, 2015; Veillette, 2009). Regulation (EU) 2019/1384 (2019) formally introduced the requirement for a third crew member (test engineer or pilot) for the purposes of coordinating complex check flight operations, referred to as an *observer* in this study. Finally, additional constructs were added from the plan and prepare framework, including *maintenance error*, the type of *test schedule* used by the airline, the planning involved in *test point entry*, and the applicable *regulatory framework* (Airbus, 2015; FSF, 2011a).

Cross-Case Analysis. Coupled with within-case analysis, the method requires searching for cross-case patterns that form the basis of a new grounded theory (Eisenhardt, 1989, 2021). The study employed pairwise comparisons that listed the similarities and differences between each selected pair. The forced comparisons included matched pairs and opposites to achieve a better understanding of potential themes emerging from case data. The cross-case analysis was conducted at three levels: i) airliner generations represented by individual cases, ii) the network of adaptive organisational units involved in check flight incidents, and iii) systemic causal propagation traced through general pathways.

Emerging Theory. The next step focused on revealing tentative themes and relationships between variables from tentative concepts, within-case analysis results, and cross-case pairwise comparisons. The central idea is to systematically compare the emerging theoretical frame with the evidence from each case. The aim is to compare theory and case data, progressively iterating toward a new theory that closely matches the evidence (Eisenhardt, 1989).

It is important to highlight that in this study the unit of analysis is not the same for within-case analyses and cross-case comparisons. For the purposes of theory building, a proposed construct or category can either be retained at an adaptive unit level, or categories can be grouped into a causal pathway. The process involved multiple iterations between case stories and the emerging accident model (theory).

A second step in shaping the theory is verifying that emerging relationships match the evidence obtained from case data. The underlying logic is replication, the logic of treating cases as a series of experiments (Yin, 2014). The idea is to confirm or disconfirm propositions in each case. Cases that confirm the proposed relationship between variables increase the validity of the study. Cases that disconfirm the proposed causal dynamics offer an opportunity to refine the theory (Eisenhardt, 1989). The study introduced additional cases for the purposes of replicating proposed constructs and categories. Extended case selection followed the same theoretical sampling plan as described earlier. The size of the extended sample is not prescribed by the method, it is primarily driven by the concept of theoretical saturation as explained in the next section.

Reaching Closure. Eisenhardt (1989) suggests two main closure issues: a) reaching theoretical saturation, and b) when to stop iterating between theory and data. The two issues are tightly coupled. Glaser and Strauss (1967) define theoretical saturation as the point where incremental learning is minimal. In a similar vein, Eisenhardt (1989) recommends stopping the iteration when the incremental improvement to the theory is minimal.

The study did not plan the number of case studies in advance. Additional cases were added to address the need to replicate proposed constructs and categories and to achieve a good fit between patterns and underlying theoretical arguments. While the number of cases is not inherent, Eisenhardt (2021, p. 153) states that case numbers between 4 and 10 are common and appear to work well for generating new theory. Chapter eight presents the extended cases and categories in tabular and narrative formats.

3.6 Reliability and Validity Threats

Yin (2014) observes that *reliability, construct validity, and external and internal validity* are the main challenges for case study research. Credibility is a similar concern for the systematic review output. This section revisits each challenge and the mitigations applied during data collection and analysis.

3.6.1 Credibility

Boeing's (2022) annual Statistical Summary of Commercial Jet Airplane Accidents is the only available literature source which contains some historical accident data about non-routine flight operations. The data is limited to 60- and 10-year totals for maintenance test flight, ferry, positioning, training, and demonstration accidents worldwide. The total counters are broken down into the number of fatal accidents, hull losses, and the number of onboard and external fatalities.

The study utilizes a well-documented and robust systematic review protocol for gathering research data from investigation report archives. To provide additional assurances that the list of non-routine events is a representative sample, the study employed a simple data triangulation method by comparing Boeing's (2022) latest accident statistics with NRFO entries in the catalogue.

3.6.2 Model Reliability

The selected multiple case method permits the induction of more reliable models than single case studies. The method embeds Yin’s (2014) replication logic. Each case can either confirm or disconfirm tentative constructs and propositions put forward through successive iterations (Eisenhardt, 1989).

3.6.3 Construct Validity

Construct validity is established by Eisenhardt’s (1989) clearly prescribed operational process. It is further enhanced by selecting multiple sources and ample evidence provided in research notes that ensure a trail for replicating the theory building process.

3.6.4 External Validity

Lunenburg and Irby (2008) define external validity as the extent to which the findings can be generalised, bound by the definition of the context in which it can be generalised. In this study, external validity is addressed by the multiple case design, as all cases are airline post-maintenance check flight accidents or incidents with the same operational objective. It follows that the findings can be extended within the context of airline MCF operations.

3.6.5 Internal Validity

Internal validity addresses the truth value of findings, as opposed to false relationships (Lunenburg & Irby, 2008). The study addressed internal validity by establishing causal relationships through pattern matching, as described in the data analysis sections in this chapter (Eisenhardt, 1989, 2021).

3.7 Summary

This chapter started with restating the research purpose and research questions and outlined the reasons for selecting a qualitative research framework for the study. The research design involves a combination of methods to test the research questions. A systematic review generates the missing research data. The initial data analysis utilises text analysis and causal mapping techniques to reveal knowledge constructs from investigation reports. Finally, a multiple case study method was selected to identify common patterns in check flight accident and incident causation, generating a new accident theory grounded in case study evidence.

CHAPTER 4 – INITIAL RESULTS

4.1 Introduction

The study intended to investigate under what conditions, and through what causal paths, maintenance check flight accidents unfold. The purpose of the study was achieved by locating check flight investigation reports in government published safety databases and examining the reports for causal descriptions. This chapter presents the research data and the initial data analysis for the first research question.

Research Question 1

What causes and contributing factors are annotated in relevant check flight investigation reports?

4.2 Historical Records

4.2.1 The NRFO Catalogue

During the first and second iteration of the systematic search, more than 52,500 accident and incident records matched the {1 Jan 1988 to 1 Mar 2020} occurrence date range in the selected occurrence databases. In accordance with the combined inclusion and exclusion criteria, 1,238 investigation records required a preliminary assessment. Having excluded a further 1,148 investigation records for routine flights, 90 non-routine occurrences were added to the catalogue. In March 2022, the database was refreshed to cover the missing {1 Mar 2020 to 1 Mar 2022} date range. This third iteration identified an additional 11 non-routine occurrences (see Table 1).

Table 1

Systematic Review Results (NRFO records)

	FIRST ITERATION						
Database	NTSB	ATSB	TAIC	BEA	BFU	AAIB	SUBTOTALS
Initial search	49,650	154	424	10	108	149	50,495
For review	178	154	2	10	1	149	494
Results	36	4	1	4	1	18	64
	SECOND ITERATION						
Database	AvH	ASN	FSS				SUBTOTALS
Initial search	183	1858	21				2,062
For review	183	540	21				744
Results	12	11	3				26
	THIRD ITERATION						TOTALS
Database	AvH	ASN	BFU	AAIB	B3A		52,636
For review	23	47	7	1	1		1,317
Results	5	3	1	1	1		101

In total, the systematic review returned 101 NRFO records which were reported between 1988 and early 2022. The NRFO catalogue includes 36 non-routine accidents and 65 incidents. Due to size limitations, it is not feasible to embed an extended version of the dataset in the thesis document. Appendix B provides an extract from the catalogue, highlighting key characteristics for each non-routine event.

4.2.2 Check Flight Accidents and Incidents

Assembling a representative catalogue of major non-routine events was the initial step in better defining the non-routine safety problem. Once the NRFO catalogue was broken down into various non-routine flight operation types, the database was filtered for MCF events. The review identified 7 check flight accidents and 15 incidents for the Western-built commercial airliner fleet (see Table 2).

A01 – Apr 1993 – DC-9-15, Margarita Island (Caracas), Venezuela: The aircraft departed Caracas on a post-maintenance test flight, carrying eight engineers and three crew members. Twenty-eight minutes into the flight the crew started the test program. A few minutes later, the pilot declared a brief MAYDAY, the aircraft entered an uncontrolled descent and crashed into the sea 16 km off Margarita Island. The aircraft disintegrated on impact and sank to a significant depth. The wreckage was not recovered. To this day, the 1993 crash remains the worst check flight accident in terms of fatal injuries (B3A, 2019).

Table 2

Maintenance Check Flight Accidents and Incidents (1988-2021)

ID	EVENT DATE	LOCATION	COUNTRY	MODEL
A01	02-Apr-1993	Margarita Island *	Venezuela	DC-9-15
I01	29-Apr-1994	Heathrow Airport	United Kingdom	Concorde
I02	22-Oct-1995	Bournemouth	United Kingdom	B737-236
I03	29-Oct-1995	San Francisco, CA	United States	B737-500
A02	22-Dec-1996	Narrows, VA	United States	DC-8-63F
I04	19-Dec-1997	Shannon	Ireland	MD-82
I05	12-May-2000	Dublin	Ireland	B747-212B
I06	25-Nov-2000	Newark, NJ	United States	MD-11
A03	08-Nov-2002	Salamanca	Spain	A340-313
I07	03-Dec-2002	Munich	Germany	A300-600
I08	11-Mar-2004	Fort Lauderdale, FL	United States	A300F4-605R
I09	01-Nov-2005	Not recorded	United Kingdom	B737-36N
I10	22-Oct-2006	London Stansted	United Kingdom	B757-204
I11	21-Nov-2007	South of France	France	A330-202
A04	27-Nov-2008	Perpignan	France	A320-232
I12	12-Jan-2009	Norwich	United Kingdom	B737-73V
I13	29-May-2009	London	United Kingdom	B737-73V
I14	07-Aug-2012	North Sea	United Kingdom	B757-2K2
A05	06-Dec-2013	Tripoli	Libya	E170
A06	20-Nov-2014	Dallas, TX	United States	B737-7H4
A07	11-Nov-2018	Alverca **	Portugal	E190
I15	13-Jul-2021	Luton	United Kingdom	A319-111

Notes: * No investigation report available. **Validation during ferry flight.

I01 – Apr 1994 – Concorde, Heathrow, United Kingdom: During a post-maintenance test flight, when the aircraft was cruising at FL570 and Mach 2.0, two outer window panels shattered in the cabin. There was no loss of cabin pressure. On return to London Heathrow, it was discovered that a third outer panel had failed immediately forward of the other two windows and the outer ply of the two-ply inner panel had also cracked at the 23 Left position (AAIB, 1994).

I02 – Oct 1995 – B737-200, Bournemouth, United Kingdom: During a post-maintenance test flight, when the aircraft was in straight and level flight at FL200 with an IAS of 290 kt, the aircraft experienced roll and yaw oscillations. According to FDR data, the autopilot and auto-throttle were disengaged, and the commander reported that the yaw damper was switched off, but the crew were unable to stop the oscillations. After a MAYDAY call, a descent was made to around FL75 and as the airspeed was allowed to reduce towards 250 kt, the oscillations began to decay rapidly and stopped. The total duration of the roll and yaw event was about seven minutes. The investigation uncovered fluid ingress in the yaw damper coupler, which resulted in a forced Dutch Roll mode of the aircraft. The location of the avionics bay, below the entry door, forward galley and lavatory area, made it vulnerable to fluid ingress. The crew actions immediately following the onset of the Dutch Roll mode did not disengage the yaw damper system (AAIB, 1998).

I03 – Oct 1995 – B737-500, San Francisco, United States: The airplane experienced an upset during a maintenance test flight while on an Instrument Landing System (ILS) approach to San Francisco International Airport (SFO). The upset occurred when the flight crew was given a go-around by approach control and the captain selected the takeoff and go-around switch. The airplane pitched up to 45 degrees nose up and the stick-shaker activated. The pilots stated that they both pushed the control column forward and attempted to retrim the nose down but were unable to overcome the increasing pitch attitude. The airplane stalled, rolled slightly, and the nose dropped. After the crew recovered from the upset, the airplane made a normal landing at SFO (NTSB, 2007).

A02 – Dec 1996 – DC-8-63, Narrows, United States: A DC-8-63 freighter was destroyed during a post-modification evaluation flight (check flight), fatally injuring all three crew members and three observers on board. The NTSB determined that the probable causes of the accident were pilot error and the failure of the airline to establish a formal check flight program (NTSB, 1997).

I04 – Dec 1997 – MD-82, Shannon, Ireland: The incident airplane departed Shannon for a ferry flight to Zurich, after a heavy maintenance visit. On reaching FL 370 there was a loss of cabin pressure control. The crew attempted to regain control in manual mode, but this was unsuccessful, and commenced an emergency descent. At FL220, cabin pressurization stabilized, and the control system responded to normal control inputs. After consultation with the onboard engineers, and after system function checks, the crew performed a stepped climb. At FL310, however, they lost control of the pressurization system again and conducted a second emergency descent. The commander decided to return to Shannon for a technical investigation and rectification (AAIU, 1998).

I05 – May 2000 – B747-200, Dublin, Ireland: The aircraft took off from Dublin Airport for a check flight. After take-off, a significant airframe vibration was encountered. The crew deduced that both airspeed indicators were under-reading significantly. Following declaration of an emergency, the aircraft returned safely to Dublin. The flap system had suffered damage during the flight. The investigation found that the static drain ports in the avionics bay, connected to both pilots' instruments, were left open (without blanking caps) after maintenance. This resulted in both airspeed indicators under-reading by a significant amount (AAIU, 2004).

I06 – Nov 2000 – MD-11, Newark, United States: During a routine sector, the airplane experienced pitch oscillations while climbing through 26,000 feet. The pilots were not injured and returned safely to Newark. After extensive troubleshooting and defect rectification, the airplane was test flown. During the test flight, oscillations were again encountered. The left inboard elevator parallel engage solenoid shut off valve was then replaced, and no further discrepancies were noted with the system (NTSB, 2001).

A03 - Nov 2002 – A340-313, Salamanca, Spain: During a post-maintenance check flight, the commander decided to test the don’t sink warning of the GPWS system and requested ATC approval to conduct two flypasts at a low height. When on final approach, the aircraft started deviating to the left of the runway axis. The aircraft continued descending while the deviation was increasing. The left and right main landing gears touched down on the apron and parts of the right outboard flap collided with an 11 m high sentry box adjacent to the airport fence (Civil Aviation Accident and Incident Investigation Commission [CIAIAC], 2003).

I07 – Dec 2002 – A300-600, Munich, Germany: Unknown to the pilots, the airplane was released by maintenance for an operational check flight. While climbing with autopilot engaged the crew noticed that the allowed airspeed would be exceeded. Initially they selected a lower speed and a higher climb rate on the autopilot panel, then the autopilot was disengaged when the airspeed increased further, and the nose started to drop. The maximum allowed airspeed was exceeded by 16 kt. A great amount of control forces had to be applied until the trim could be corrected by electrical trim. A cabin crew member suffered minor injuries when the original flight attitude was restored. The flight was continued with disengaged autopilot and no further incidents. The investigation found that a combination of design and maintenance errors led to the serious incident (BFU, 2004).

I08 – Mar 2004 – A300F4, Fort Lauderdale, United States: During an operational check flight, the airplane had a failure of all eight main landing gear tires during the landing roll. The tower controller then informed the crew that emergency equipment was rolling because of visible smoke and fire. The captain performed the emergency evacuation checklist, and the pilots exited the airplane via an emergency slide. Prior to departure, the flight crew were given paperwork to perform an alternate brake system test in-flight. The investigation found that the crew inadvertently set the parking brake during the system test, a scenario which was confirmed by an Airbus test flight (NTSB, 2005).

I09 – Nov 2005 – B737-300, United Kingdom: The crew experienced a large, uncommanded pitch-up reaction during a manual reversion test when flight control system hydraulics were switched off at FL350. The cause of the incident was identified as an incorrect adjustment of the elevator balance tabs. The pilot rolled the aircraft through to 65° before releasing the controls, restoring the hydraulics, and re-establishing control. The crew repeated the test, the second time managing to control the pitch reaction, before returning the airplane to the maintenance base (AAIB, 2010b).

I10 – Oct 2006 – B757-200, Stansted, United Kingdom: Shortly after reaching cruise altitude on a scheduled passenger flight, a blue haze was observed in the passenger cabin. A precautionary diversion was made to London Stansted, where an emergency evacuation was carried out. The blue haze could not be reproduced by maintenance, which included engine ground runs. A planned post-maintenance check flight was aborted during takeoff when smoke entered the flight deck and the cabin. Further investigation, which included ground runs at a higher engine power, identified the source of the smoke to be the right-hand engine (AAIB, 2009).

I11 – Nov 2007 – A330-200, Southern France: Airbus was performing a customer acceptance flight with 2 crew and 10 engineers on board, when the cabin pressure was rapidly lost while cruising at FL410. The cabin altitude reached FL300. The flight crew donned their oxygen masks and initiated an emergency descent. Two engineers in the cockpit and one in the cabin lost consciousness, then later recovered after they had been administered oxygen by other occupants. The crew returned to Toulouse for a safe landing. The investigation found that a cabin pressure controller had failed. There was less safety equipment on board than occupants and the passenger oxygen mask containers had been mechanically locked by Airbus, so that the masks did not have to be repacked after scheduled tests. The crew did not transmit a distress message and did not set the transponder to 7700 during the serious incident (BEA, 2008).

A04 - Nov 2008 – A320-200, Perpignan, France: The airplane operated a post-maintenance check flight, in the context of ending a lease agreement, when it was destroyed upon impact off the coast of Canet-Plage. This major accident involved two airlines and several aviation authorities and safety boards worldwide. The Perpignan tragedy created significant interest and concern within the airline industry and the final BEA investigation report played a key role in shaping the current safety response (BEA, 2010).

I12 – Jan 2009 – B737-700, Norwich, United Kingdom: The airplane operated a combined check flight and customer demonstration program, having just completed a maintenance visit, when it experienced a serious in-flight upset and loss of control incident. The airplane violently pitched down and lost approximately 9,000ft before the pilot was able to recover and landed it safely. This serious incident served as another major wake up call to the industry, confirming that the check flight safety problem is not limited to a particular manufacturer or design philosophy (AAIB, 2010b).

I13 – May 2012 – B737-700, London, United Kingdom: The same airline operator (involved in I09, I12, and I13) experienced a large pitch down and moderate roll when conducting an asymmetrical flight control forces check during a combined check and demonstration flight. Prior to the flight, there was confusion between the three management pilots regarding the in-flight test procedure. During the test, several required steps were unintentionally missed. After the first pitch down event, the crew opted to use another test procedure, and by doing so, they again missed required steps that left the rudder boost unpowered. Without hydraulic assistance for the rudder, they could not correct the roll during the second pitch down event. The commander of the previous incident (I12) acted as copilot during the May 2012 event (AAIB, 2010b).

I14 – Aug 2012 – B757-200, North Sea, United Kingdom: During a post-maintenance test flight, the crew identified a lateral fuel imbalance, determined that fuel was leaking from the right engine, then shut down the engine. The crew made a single-engine diversion and landed at Newcastle Airport without further incident. The investigation found that the fuel leak was caused by a damaged O-ring seal that had been recently replaced due to embodiment of a service bulletin (AAIB, 2013).

A05 – Dec 2013 – E170, Tripoli, Libya: The aircraft departed for a check flight to deploy and test the Air Driven Generator (ADG). The aircraft initially climbed to 7000 feet, then descended to 3000 feet to begin the ADG test. The two integrated drive generators and the batteries were disconnected, the ADG deployed properly, and the engineer confirmed the test. The aircraft continued to operate in this configuration with the ADG remaining the only electrical power source. The aircraft touched down and the ADG ceased providing electrical power, causing the aircraft to veer off the runway. There were no injuries, but the aircraft sustained substantial damage (Libyan Civil Aviation Authority, 2020).

A06 – Nov 2014 – B737-700, Dallas, United States: The airplane was taxiing for a functional check flight with two pilots onboard. Another company airplane was holding on a taxi lane, operating a scheduled domestic flight with 177 passengers and crew members onboard. When the two airplanes collided, one suffered minor damage to the wingtip, while the other sustained major damage to the horizontal stabilizer. There were no injuries to the occupants of either airplane (NTSB, 2015).

A07 – Nov 2018 – E190, Alverca, Portugal: The airplane departed for a post-maintenance validation and ferry flight. Immediately after take-off, in adverse weather conditions, the crew realized that they had no effective control of the airplane and declared an emergency, while trying to diagnose the abnormal aircraft attitude. After two non-stabilized approaches, they managed to land on a third attempt. One crew member suffered minor injuries and the airplane was written off. The investigation found that the aileron cables were incorrectly installed during maintenance, which resulted in both ailerons acting in the opposite direction of control yoke commands (GPIAAF, 2020).

I15 – Jul 2021 – A319-100, London Luton, United Kingdom: During a post-storage check flight, the aircraft carried out a high speed rejected takeoff above V_1. The investigation found that there was a major discrepancy between the commander and first officer's airspeed indications. The unreliable airspeed indication occurred because of a blockage in the pitot tube system following a long storage period on the ground. The pilots were aware of the potential for unreliable airspeed due to an increase in similar events reported within their airline and around the world during the global pandemic (AAIB, 2022).

In terms of safety loss, the seven check flight accidents identified by the review amount to 1 in 5 non-routine accidents listed in Appendix B. Even more concerning is the fact that three of those check flight accidents (A01 Margarita Island, A02 Narrows, and A04 Perpignan) led to multiple fatalities. As a result, 24 lives were lost between 1993 and 2008. In addition to those fatal accidents, two more airplanes were destroyed during check flying, a considerable portion of hull losses recorded in the review period.

4.2.3 Data Triangulation

Table 3 presents a comparison between data generated by the systematic review and Boeing's (2022) non-routine accident statistics. The table mirrors the structure of Boeing data. Rows correspond to 10-

year operational periods, while columns reflect counters for accidents, hull losses, and onboard fatalities.

For the operational period between 1991 and 2000, the NRFO catalogue matches Boeing (2001) accident and hull loss data, although the number of onboard fatalities identified by the study is almost double of that reported by Boeing. In part, the discrepancy can be traced to a DC-8 training flight accident in 1996 that claimed 22 lives (refer Appendix B). Boeing data appears to reflect the number of operating crew members in the cockpit, not all onboard fatalities.

For the period between 2001 and 2010, Boeing (2011) reports four more hull loss accidents than this research study. The number of fatal accidents and onboard fatalities are identical. Boeing's (2022) hull loss definition includes any aircraft that is damaged and not repaired. There is no equivalent loss category in government safety databases which may explain the discrepancy.

Table 3

Non-routine Accidents in Catalogue v. Boeing Statistics

	Boeing statistics			NRFO Catalogue		
Years	Accidents	Fatal	Hull-loss	Accidents	Fatal	Hull-loss
2011-20	6	0	4	9	0	4
2001-10	11	3 (17)	8	7	3 (17)	4
1991-00	15	10 (34)		15	10 (66)	

Note: Numbers in brackets indicate onboard fatalities recorded within the reporting period.

For the period between 2011 and 2020, the NRFO catalogue includes three more accidents than Boeing (2021) data. Accident records in the catalogue were validated through at least two independent data sources during the review. The systematic review's inclusion criteria are identical to Boeing's selection rules, and as such, it is plausible that some non-routine events were incorrectly classified in Boeing's (2021) source data as routine commercial flights.

As outlined in the methodology section, the initial part of the causal analysis process was aimed at understanding what causal or contributing factors were annotated in check flight investigation reports. Other than the 1993 DC-9 accident at Venezuela, an ICAO template investigation report was readily available from a primary or secondary data source selected for the systematic review. For the 2007 A330 incident in France and the 2013 E170 accident in Libya, no official English translation was published by the investigating safety agency. The Aviation Herald database contains unofficial English summaries for both events. The unofficial versions enabled the use of 21 check flight investigation records for the purposes of the research study. For the pre-1991 years, the catalogue contains entries for every known major non-routine accident, including the 1988 Air France A320 tragedy at Mulhouse-Habsheim, the 1990 Faucett B727 ferry flight which claimed sixteen lives, and the 1990 Omega B707 ferry flight at Marana which ended in one fatality and two serious injuries. By extension, the study assumes that accident records in the NRFO catalogue are representative for the complete Jan 1988 to Mar 2022 period.

4.3 Data Analysis

4.3.1 Text Analysis

The investigation reports cited in the previous section are available to the reader, and as such, only two examples are embedded here to illustrate the results of the text analysis step. When full causal sentences are extracted from the report, it is important to capture both the structure and the temporal logic in the causal argument (see Figure 5 and 6). For ease of reference, probable causes are highlighted by a shaded background and a rectangular boundary is drawn around the relevant causal nodes. Causal relations are illustrated by arrows (solid lines) between probable causes and dashed lines between contributing factors and other nodes. The order of causal nodes is determined by temporal logic, as annotated in the source investigation report.

The sharp contrast between the two examples is obvious for any observer, even without knowing the actual causal arguments. While both the Spanish and Portuguese authorities are required to identify probable causes, the structure of their causal summary is different. Figure 6 illustrates the A03 2002 A340 accident at Salamanca (CIAIAC, 2003). A03 reflects an RCA technique that focuses on probable causes and the immediate event chain prior to incurring the safety loss. Figure 7 illustrates the A07 2018 E190 accident at Alverca. A07 nominates probable causes as well, but the investigation’s focus is on systemic contributing factors. There is a 17-year gap between the two check flight investigations that aligns with evolving investigation methods (Burns, 2000; Sklet, 2002; Rasmussen, 1997; Wienen et al., 2017). This simple comparison can readily illustrate the benefit of building concept maps as visual aids, even in a rudimentary format.

Figure 6

Text Analysis for A03 2002 Salamanca

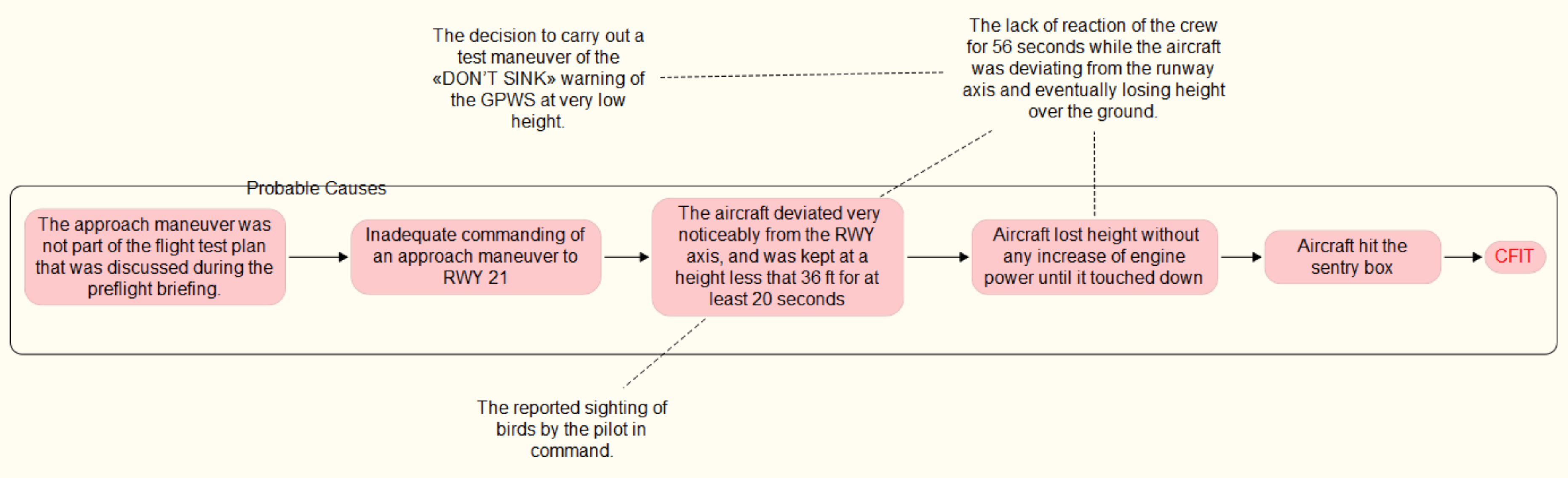
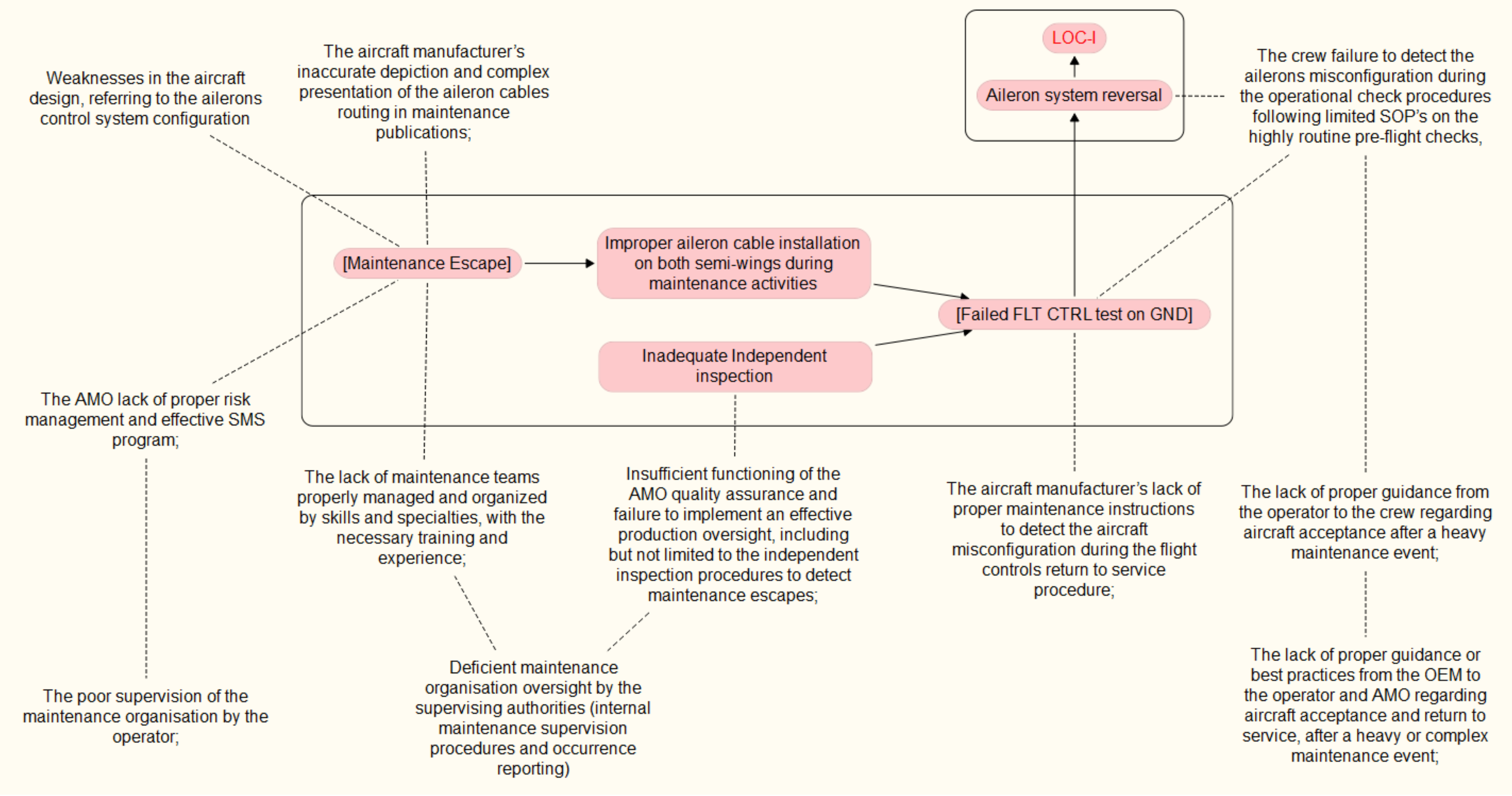


Figure 7

Text Analysis for A07 2018 Alverca



The two examples also highlight the difficulty involved in comparing causal factors cited in investigation reports which heavily rely on natural text. While a single chain of events is easy to follow, it is a very narrow description of a single causal path. A broader scope provides a richer context and more causal details, but the causal argument cannot be easily deciphered from the investigation report, even if the template conforms to ICAO recommended standards (Dismukes et al., 2007).

4.3.2 Causal Categories

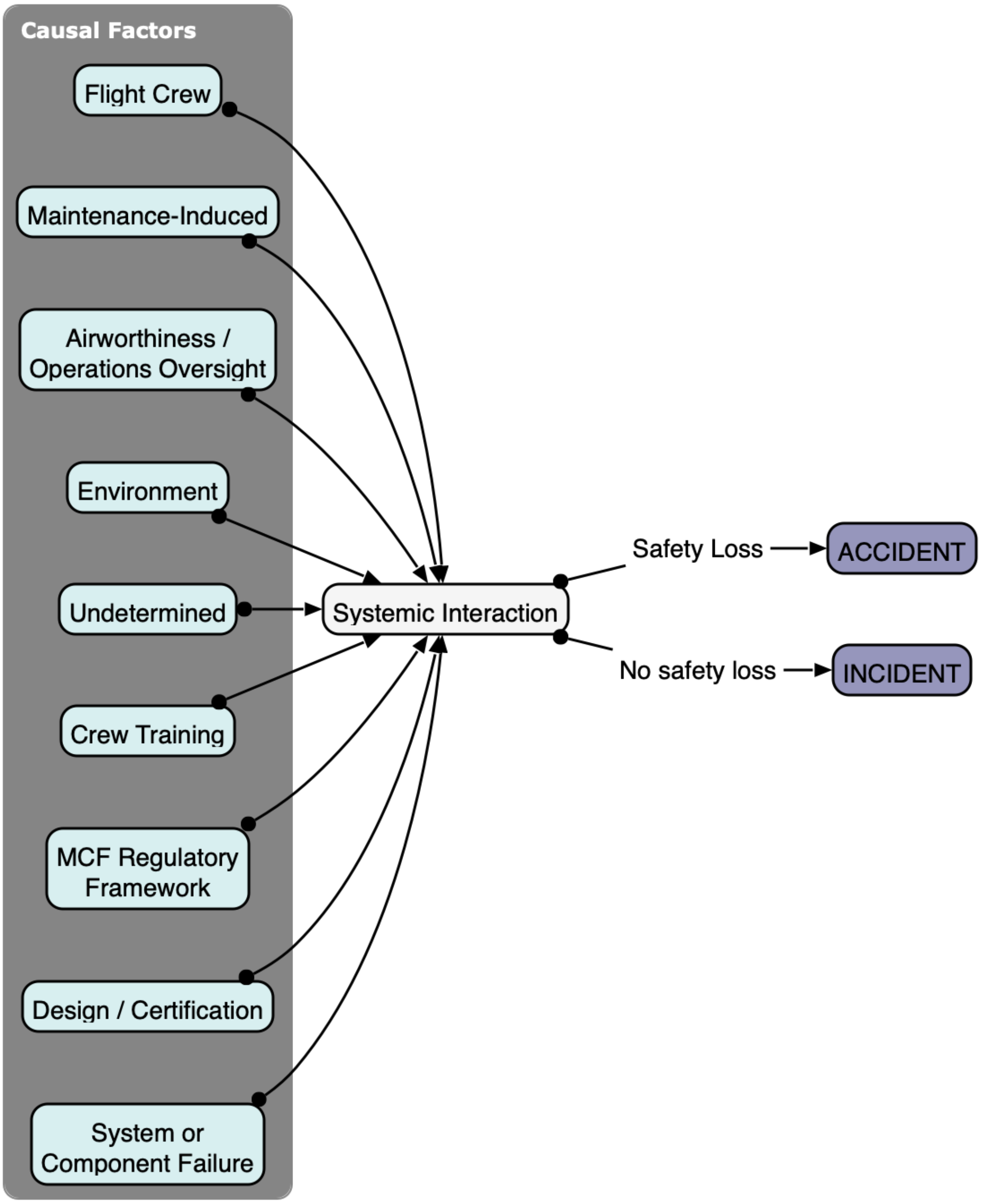
Borrowing Holloway and Johnson's (2004) text analysis method, the next step involved replacing long-form text with short-form causal labels. The classification table was embedded in Appendix C for reference. Once the short-form labels were available, the causal factors were progressively grouped into substantially similar causal categories (see Figure 8).

The study found nine primary categories across check flight investigation reports. For the purposes of clarity, the thesis uses […] square brackets whenever the text refers to a causal category, as opposed to individual causal or contributing factors:

1. [System or Component] group, including any system, component, functional failure or malfunction.

Figure 8

Causal Categories in Check Flight Reports (1988-2021)



2. [MCF Regulatory framework], including any causes or factors related to missing or unclear regulations, guidance, or shortfalls in regulatory oversight.
3. [Crew Training] including shortfalls identified in crew training or in simulation devices.
4. [Environment] including all causal factors induced by the operating environment.
5. [Airworthiness / Operations] including systemic factors induced by shortfalls in airline operational management and oversight.
6. [Maintenance] causal group, including any systemic or specific maintenance issues, or maintenance error labels annotated in check flight investigation reports.
7. [Flight Crew] including any causes, factors, or error labels assigned to operating crew members and their recovery attempts.
8. [Undetermined] reasons when the safety investigation report did not identify causal or contributing factors.

At this stage of the analysis, the two most frequent causal factors annotated in the reports were *incorrect pilot action* and *component failure* allocated to the [Flight Crew] and [System or Component] categories, respectively.

Figure 8 goes further than simply providing a graphical summary of the primary causal categories involved. It introduces causal maps to the analysis framework. The nine categories are visible on the left-hand side of the diagram. The boxes on the right indicate the safety loss, depending on the interactions between the various systemic causes and contributing factors. In turn, that systemic interaction is illustrated by the individual arrows pointing at a central causal node, which refers to the concatenation of events induced by individual systemic causes in relevant causal categories.

4.3.3 Comparison with NTSB Study

Holloway and Johnson's (2004) study describes an independent analysis of primary and contributory causes annotated in 26 major NTSB aviation accident reports issued between 1996 and 2003. While the only non-routine example in the NTSB sample is the 1996 DC-8 accident (A02) at Narrows, a comparison between the two studies can provide an insight into main causal categories annotated in routine and non-routine investigation reports.

The NTSB sample revealed 10 causal categories. Seven categories are common with this study, including causes induced by maintenance, airline operations, regulations, equipment failure, aircraft design, the environment, or undetermined reasons. Their human error group appears to be very similar to the [Flight Crew] category described in this study, except for ATM failure and manufacturing defects which did not appear in check flight investigation reports as causal factors. Contrary to the NTSB sample, human error labels remain more prevalent in check flight investigation findings. At least one form of human error is cited in 14 examples of the check flight sample. The NTSB had moved away from a traditional view on human error in the early 1990s, which was reflected in the NTSB study outcome (Baker et al., 2008; Holloway & Johnson, 2004; Strauch, 2002, 2017).

4.3.4 Causal Logic

The next step of the initial data analysis phase is to draw trivalent causal maps that illustrate causal arguments at a primary category level. In total, 21 trivalent causal maps were built for this study. The causal maps are embedded in the thesis for the reader’s reference (see Figure 9 to 29). As highlighted in the previous chapter, while these diagrams reflect the causal arguments embedded in the relevant investigation report, they cannot be considered as complete accident models.

As an example, the following paragraphs walk through the trivalent map built for the 2002 A340 accident (A03) at Salamanca to demonstrate the causal links between investigation findings and the causal graph (see Figure 17). The CIAIAC (2003) investigation found several probable causes and contributing factors. The relevant category was added in […]

brackets at the start of each paragraph, for the purposes of mapping investigation findings onto the causal map.

Probable causes (+1):

> [Flight Crew] It is considered that the most probable cause of this event was the inadequate commanding of an approach maneuver to runway 21 of Salamanca Airport, during which the aircraft deviated very noticeably from the runway axis and was kept at a height less than 36 ft for at least 20 seconds, and then lost additional height without any increase of engine power until it touched down and hit the sentry box (p. 33).
>
> [Flight Crew] This maneuver was not a part of the flight test plan that was discussed during the preflight briefing (p. 33).

Contributing factors (+0.5):

> [Flight Crew] The decision to carry out a test maneuver of the «DON'T SINK» warning of the GPWS at very low height (p. 34).
>
> [Flight Crew] The lack of reaction of the crew for 56 seconds while the aircraft was deviating from the runway axis and eventually losing height over the ground (p. 34).
>
> [Environment] The reported sighting of birds by the pilot in command (p. 34).

Figure 9

I01 – 1994 Heathrow

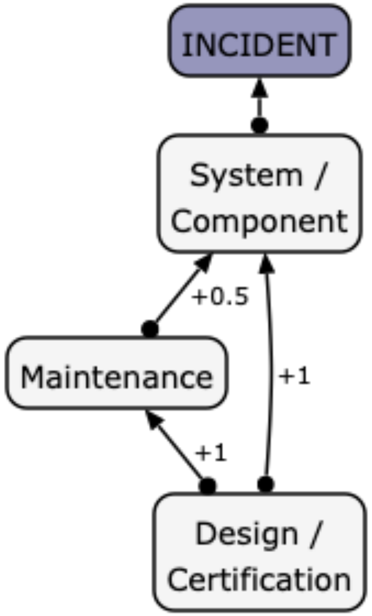


Figure 10

I02 – 1995 Bournemouth

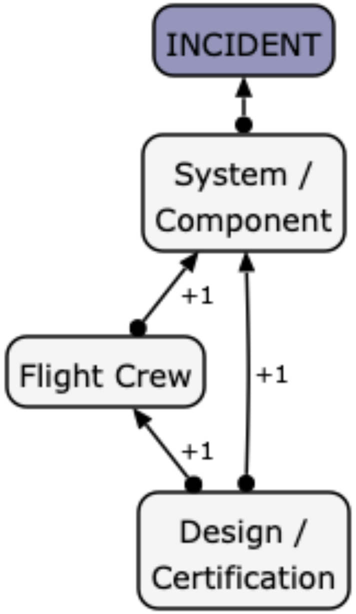


Figure 11

I03 – 1995 San Francisco

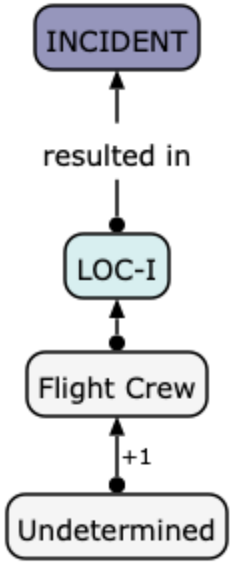


Figure 12

A02 – 1996 Narrows

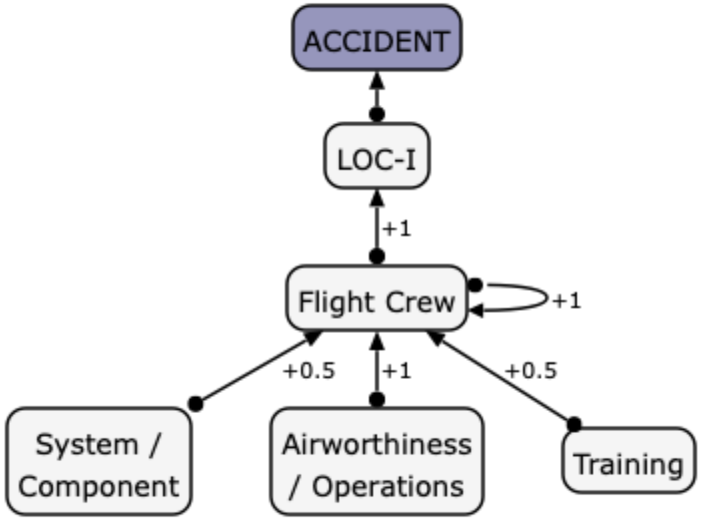


Figure 13

I04 – 1997 Shannon

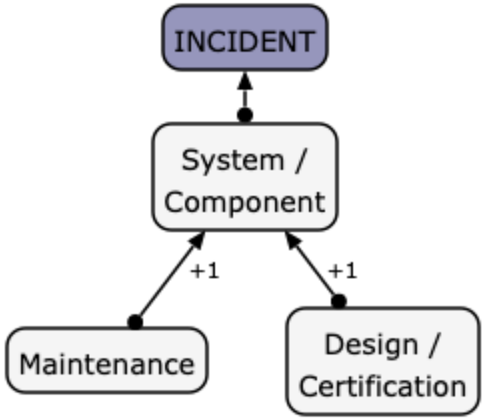


Figure 14

I05 – 2000 Dublin

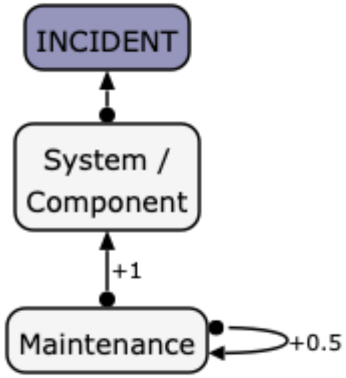


Figure 15

I06 – 2000 Newark

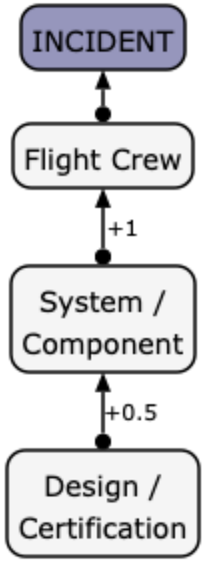


Figure 16

I07 – 2002 Munich

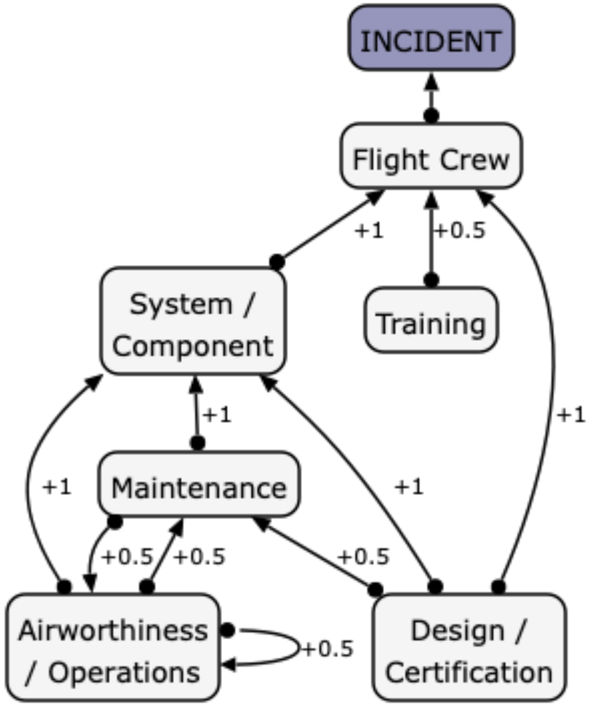


Figure 17

A03 – 2002 Salamanca

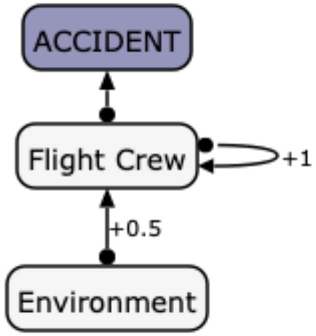


Figure 18

I08 – 2004 Fort Lauderdale

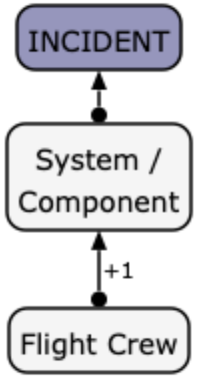


Figure 19

I09 – 2005 United Kingdom

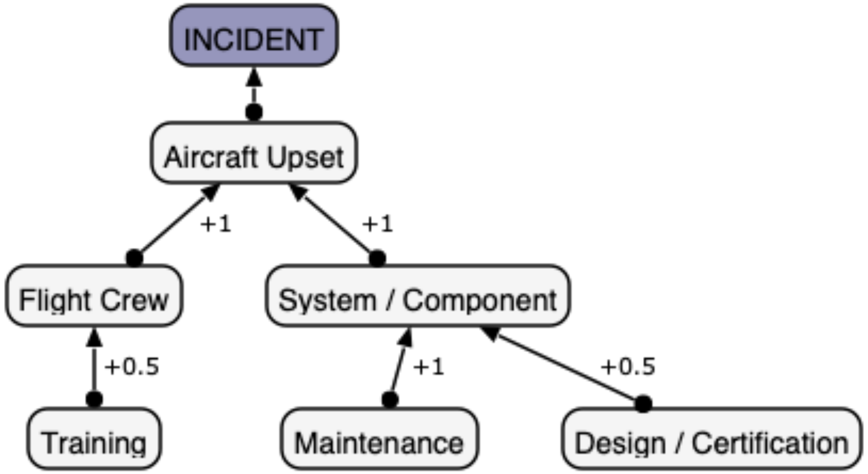


Figure 20

I10 – 2006 Stansted

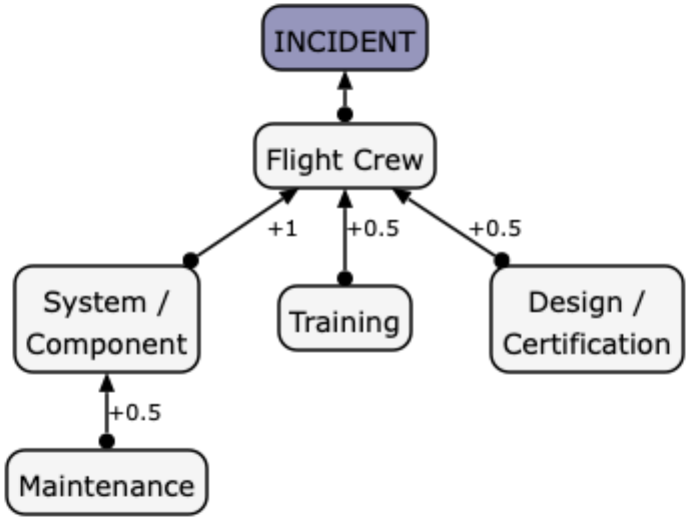


Figure 21

I11 – 2007 Southern France

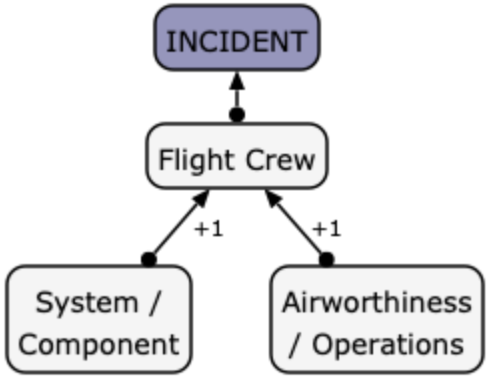


Figure 22

A04 – 2008 Perpignan

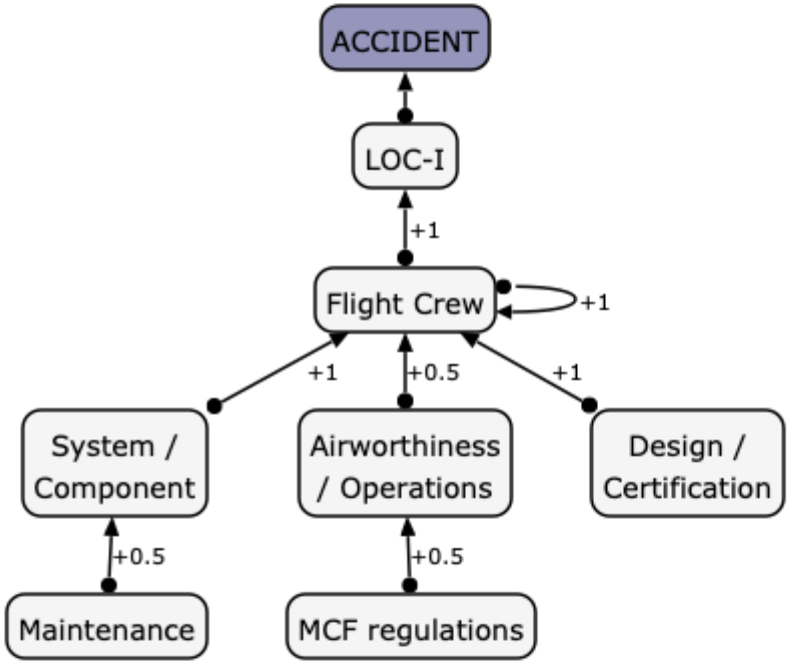


Figure 23

I12 – 2009 Norwich

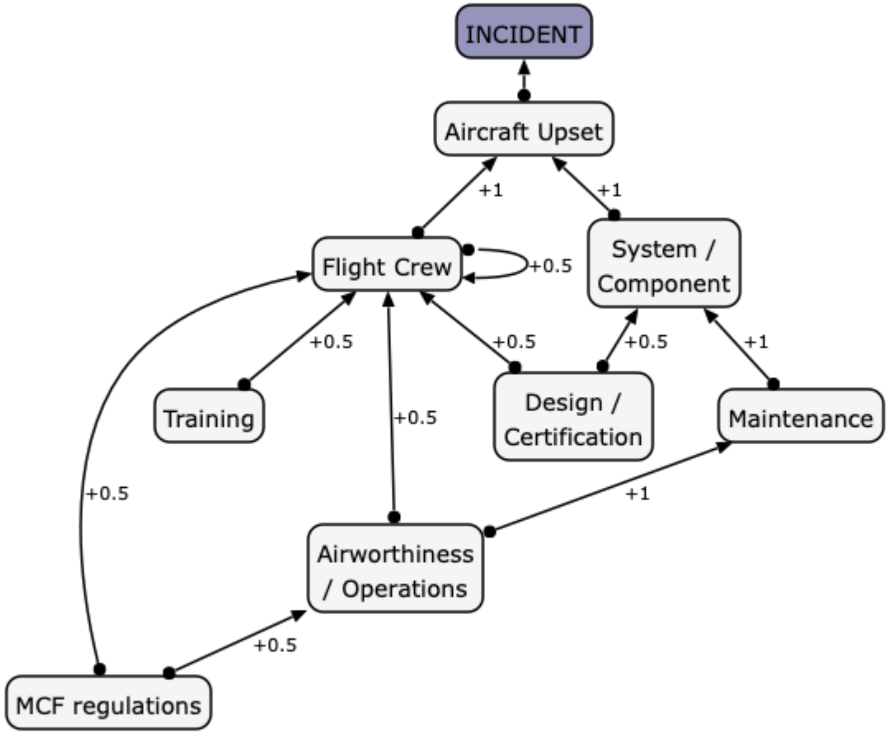


Figure 24

I13 – 2009 London

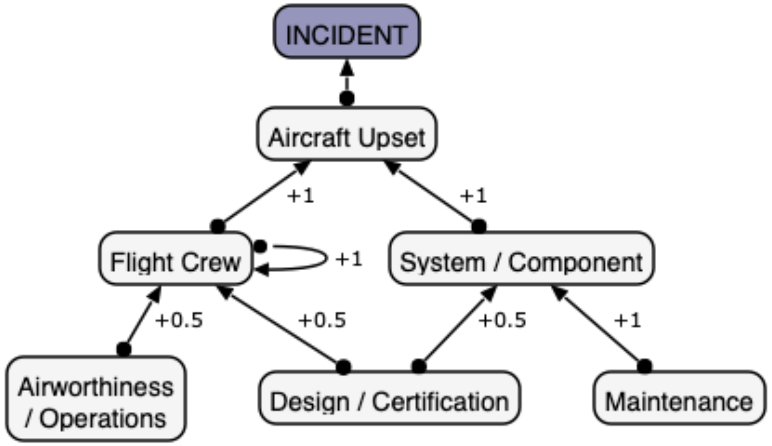


Figure 25

I14 – 2012 North Sea

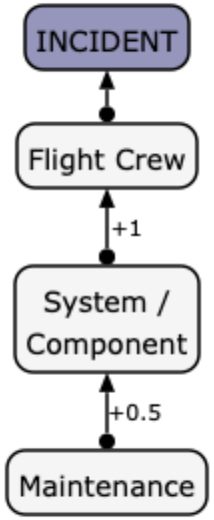


Figure 26

A05 – 2013 Tripoli

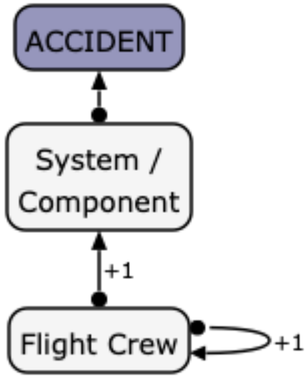


Figure 27

A06 – 2014 Dallas

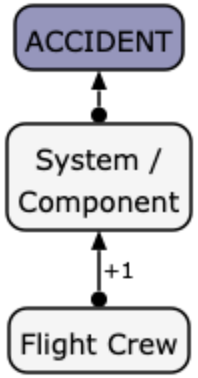


Figure 28

A07 – 2018 Alverca

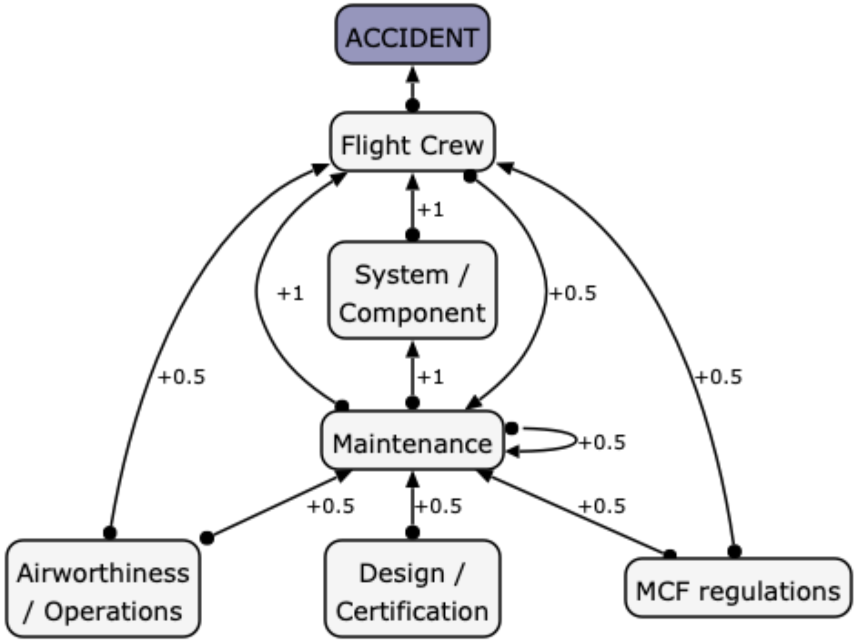
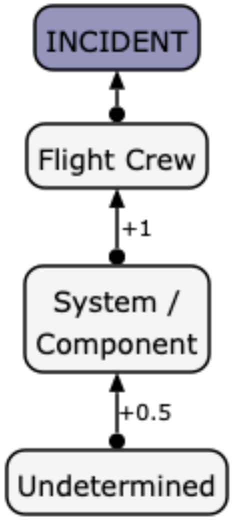


Figure 29

I15 – 2021 Luton



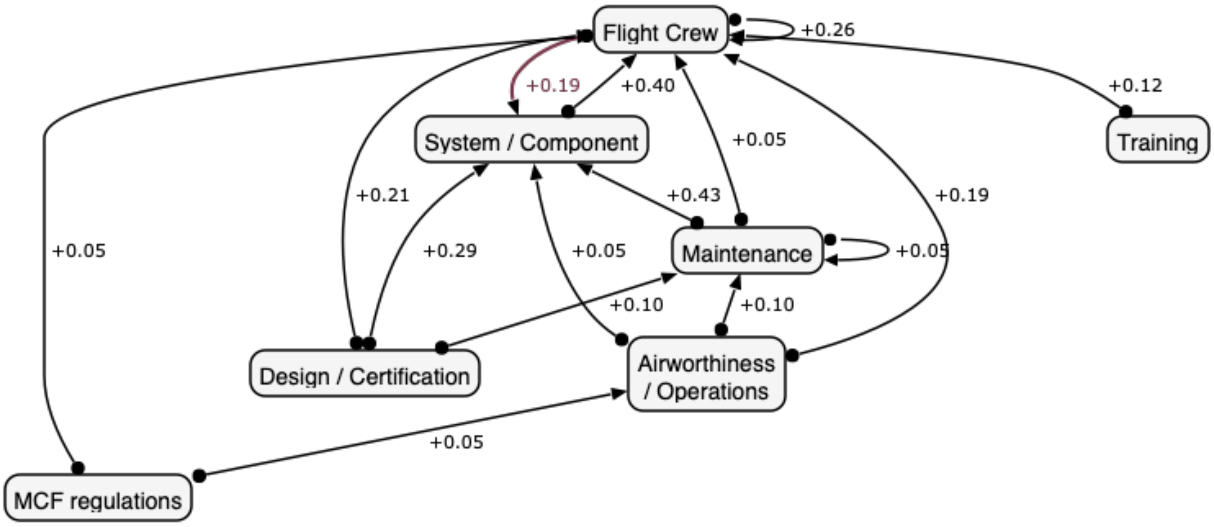
4.3.5 Knowledge Map

Once the contribution from each MCF event was accounted for by trivalent causal maps, the individual graphs were overlaid to form a single causal map, referred to as the *knowledge map* for the purposes of the research study. The fuzzy causal algebra involved in combining the trivalent maps is explained in the methodology chapter and is not repeated here.

Figure 30 illustrates the causal knowledge elicited from all check flight investigation reports included in the study. For the purposes of clarity, this version of the knowledge map only includes causal categories and links where the causal strength is at least {+0.05} between the nodes. The knowledge map contains a fuzzy output. Due to the limited sample size, the augmented connection matrices are sparse. In other words, the individual maps are extended to account for every causal category discussed in check flight reports, but not every category is mentioned in each investigation report (Kosko, 1988).

Figure 30

Knowledge Map Built from Check Flight Investigations (1988-2021)



As discussed earlier, not every single causal linkage is visible on the knowledge map. The graph represents a second-order construct where arrows reveal causal patterns between categories. The knowledge map is not intended to be a complete accident model, it is a logical construct that assembles the overall structure, temporal logic, and causal patterns annotated in check flight investigation reports.

Following the causal links in the knowledge map, a typical causal pattern for a check flight occurrence would be described as: maintenance induced error(s) may lead to critical system or component failure(s) prior to or during the check flight sequence. A critical system failure requires the flight crew to intervene in the unfolding safety loss scenario. Flight crew actions may contribute to system failures or malfunctions, while crew training, system design, and airline operational factors may impact the pilot’s ability to recognise and recover from the in-flight scenario. Design and certification shortfalls may also contribute to system or component failures, while airworthiness oversight may be a contributing factor to maintenance error(s).

There are three incident reports in the sample (records I01, I04, and I05) which annotate a causal link between the [System / Component] category and the actual safety outcome without assigning any causal factors from the [Flight Crew] category. All other incident investigations concluded that pilot action / inaction was part of the causal chain, either as an immediate cause, or as a factor contributing to system or component failure(s). In contrast, every single accident investigation (records A02; A03; A04; A05; A06, and A07) assigned one or more immediate cause(s) from the [Flight Crew] category. This reflects a tendency to affix human error labels whenever the accident flight ended in a considerable safety loss. The [Flight Crew] category is central to the overall structure and temporal patterns elicited from check flight investigation reports.

4.4 Summary

This chapter discussed the results of the data gathering and initial data analysis phases. The systematic review identified 101 non-routine occurrences worldwide for the selected 1988-2022 operational period. Accident records in the NRFO catalogue were compared to Boeing statistics to enhance the credibility of research data used in the study. In total, the review yielded 22 check flight records, including 7 accidents and 15 incidents within scope.

The initial data analysis phase revealed 9 primary causal categories that are common to check flight investigations. Investigation findings were mapped onto trivalent causal maps for each MCF event, then the individual graphs were combined into a single knowledge map. The knowledge map revealed the common causal patterns annotated in check flight investigation reports, in response to research question one.

The next three chapters will present and analyse the case studies conducted across three different airliner generations, followed by a cross-case analysis and synthesis of research findings.

Chapter 5 – CASE A: DC-8 STALL SERIES

5.1 Introduction

Event: Uncontrolled Flight Into Terrain ABX Air (Airborne Express)

Douglas DC-8-63, N827AX

Location: Narrows, Virginia – United States of America

Date: 22 December 1996

Agency: NTSB

On 22 December 1996, a Douglas DC-8-63 operated by Airborne Express was destroyed when the airplane impacted the terrain at high speed in the vicinity of Narrows, Virginia. There were three flight crew members (two pilots and a flight engineer) and three maintenance engineers on board, all of whom suffered fatal injuries upon impact. The post-modification functional evaluation flight (check flight) departed from Piedmont Triad International Airport (GSO) in North Carolina. The flight was conducted under Part 91 on an instrument flight rules (IFR) flight plan. The accident occurred during the hours of darkness.

The NTSB's (1997) final report concluded that the probable causes and factors contributing to the accident were:

> The National Transportation Safety Board determines that the probable causes of this accident were the inappropriate control

> inputs applied by the flying pilot during a stall recovery attempt, the failure of the nonflying pilot-in-command to recognize, address, and correct these inappropriate control inputs, and the failure of ABX to establish a formal functional evaluation flight program that included adequate program guidelines, requirements and pilot training for performance of these flights. Contributing to the causes of the accident were the inoperative stick shaker stall warning system and the ABX DC-8 flight training simulator's inadequate fidelity in reproducing the airplane's stall characteristics (p. v).

Figure 31 provides a graphical summary of the causal logic embedded in the section quoted above. In an earlier phase of the study, trivalent causal maps delivered a graphical representation for main causal categories extracted from check flight investigations, including the NTSB (1997) report (see Figure 12 duplicated on the next page).

As discussed in chapter three, it is inevitable that at a higher level of abstraction certain causal details remain hidden. However, when the two diagrams are compared, a simple linear translation can reveal those hidden details. For example, "inappropriate control inputs applied by the flying pilot during a stall recovery attempt" was nominated by the NTSB (1997, p. v) as one of the probable causes. That finding translates to a causal link between the [Flight Crew] and LOC-I nodes and the {+1} causal strength (probable cause) assigned to the edge between those nodes in Figure 12.

Figure 12

A02 – 1996 Narrows (duplicated for ease of reference)

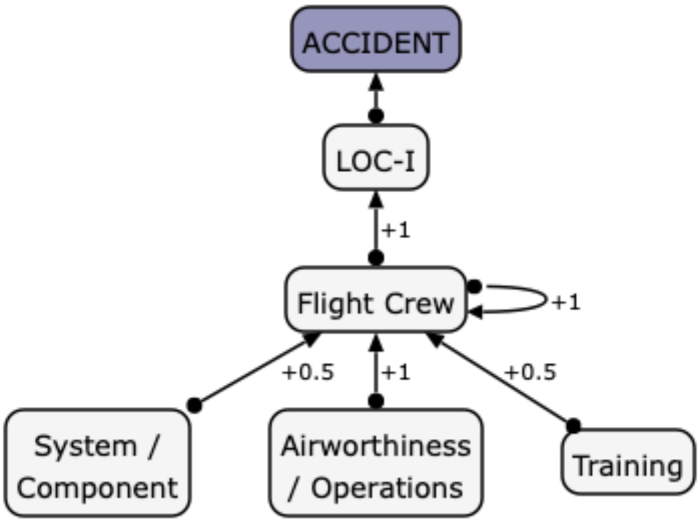
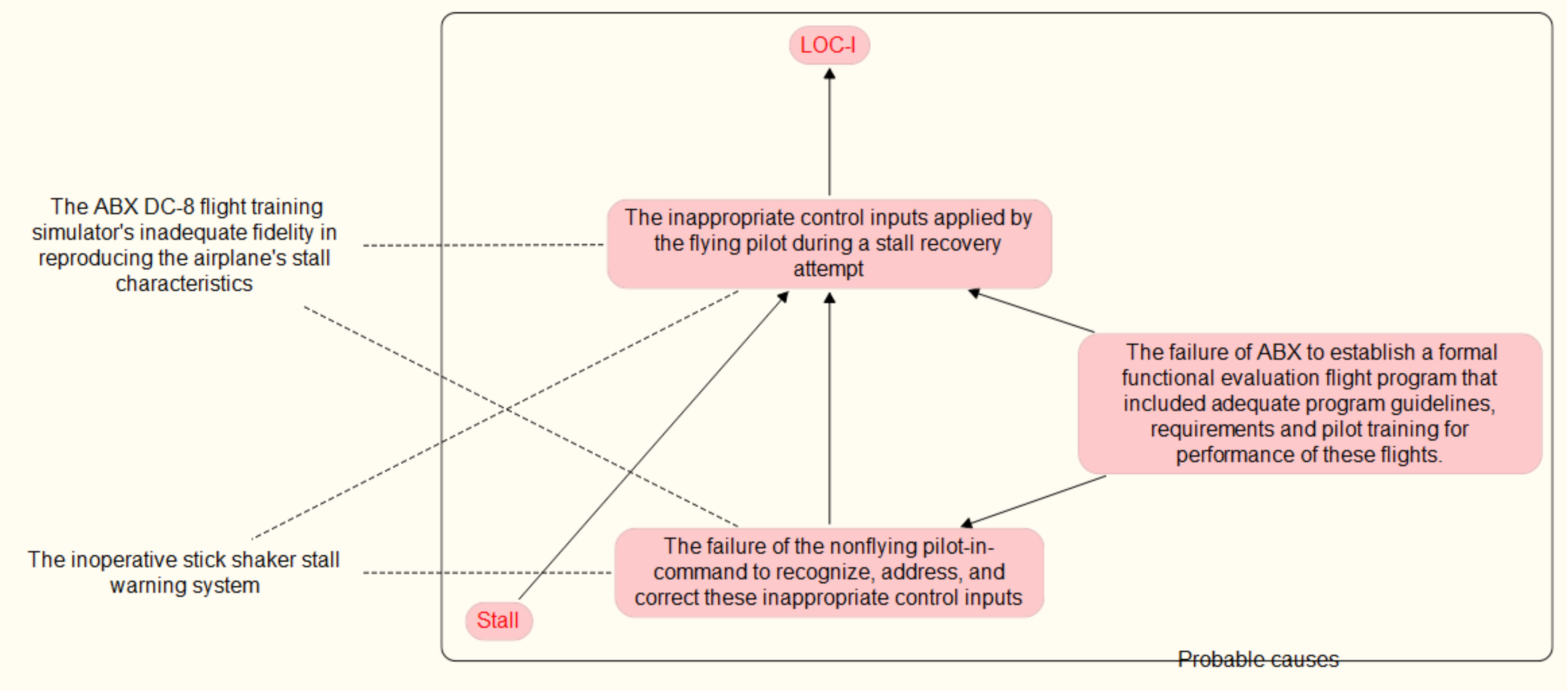


Figure 31

Text Analysis for 1996 Narrows



The Independent Safety Board Act of 1974 requires the NTSB to investigate transportation accidents, "establish the facts, circumstances and…probable cause" of each individual accident, and issue safety recommendations to prevent similar accidents (Independent Safety Board Act, 2006, § 1131). While this case study relies on the final report issued by the NTSB, it has a slightly different objective, supported by focus questions. In this instance, the case study intends to reveal a) the reasons why it made sense for the ABX crew to proceed with the stall series and b) the reasons why the recovery attempt was not successful.

5.2 Data Analysis

5.2.1 Nominated Transition Point

Based on the sequence of events described in the NTSB (1997) report, the decision to proceed with the stall series was nominated as the point where the check flight started to transition from the original test program to a recovery phase. The case study describes noteworthy events leading up to the accident flight and significant events pre- and post-transition. The pre-transition narrative focuses on factors contributing to the crew's decision to accept and start the stall series, intentionally stalling the aircraft in accordance with the scheduled test point. The post-transition narrative revisits available evidence in the context of the failed recovery attempt, including the point when the crew realized the need for recovery and initiated the attempt.

5.2.2 Noteworthy Events Prior to The Accident Flight.

The accident airplane had received a major overhaul and extensive modifications at Triad International Maintenance Corporation (TIMCO), an FAA-certified repair station, prior to entering service in ABX cargo operations. The maintenance layover included an airframe D check, a cockpit upgrade aligning the flight deck interface with the ABX fleet, installation of a cargo handling system and Stage III hush kits installed on the engines and pylons (NTSB, 1997, p. 7). This level of complex airframe, mechanical and avionics work package cannot be performed without major disassembly, which inevitably leads to flight-safety critical systems being disturbed, and as a result, extensive ground testing, adjustments and rigging required during the layover. TIMCO had to replace multiple fuselage belly skin panels due to widespread corrosion damage, which extended the ground time to six months (NTSB, 1997, p. 7).

On 21 December 1996, TIMCO released the airplane for the first check flight, but that flight was terminated by the ABX crew when a low fluid quantity caution was indicated in the cockpit for one of the main hydraulic systems. TIMCO had to reopen the maintenance package, perform troubleshooting, and repair the hydraulic system. The crew had originally scheduled a 1320K (UTC-5) departure for 22 December, however additional maintenance delays prevented maintenance release and the airplane only departed GSO at 1740K (UTC-5) (NTSB, 1997, p. 1). While the NTSB (1997, p. 49) noted in its conclusions that ABX "should have required completion of the check flight by sundown or should have established adequate night and weather limitations" due to the maintenance delays incurred during that day, the crew performed all test manoeuvres when it was dark, and they were not supported by a visible natural horizon.

5.2.3 Significant Events Leading up to the Transition Point.

Following takeoff, ATC assigned a 13,000 to 15,000 feet AMSL airspace block, in line with a "round-robin" IFR clearance along a planned route from GSO to waypoints in West Virginia, Kentucky, and Virginia, before returning to GSO. Weather records indicated that cloud tops were just below 14,000 feet along the route. CVR comments indicated that the DC-8 flew in and out of the clouds and the PF observed ice buildup once they reached the assigned block altitude. Ice buildup was clearly a concern for the PF, when he repeatedly commented "We're getting' a little bit of ice here… probably get out of this" and "We just flew out of it, let's stay here for a second.", refer CVR timestamp 1748:34 and 1752:19, respectively (NTSB, 1997, p. 72).

Having successfully performed several landing gear, hydraulic, and engine system functional checks, the flight engineer reminded the two pilots that the next test point required was the stall series, refer CVR timestamp 1805:37. ABX used an abbreviated in-flight evaluation form that required crew members to calculate and record speed values for the stall series. Following the flight engineer's reminder, a conversation ensued between the crew members about the applicable test conditions.

At 1805:56, the Pilot Monitoring (PM) shared his calculations about stopping the trim and the expected stall speed as "one eighty four [or 1.5 times the stall speed], and…we should get uh, stall at uh, one twenty two. I'm going to set that in my, interior bug." (NTSB, 1997, p. 93).

At 1806:10, the flight engineer confirmed that he expected stick shaker activation 5 percent above the calculated stall speed, when he stated, "shaker at one twenty eight if you just … call out your numbers, I'll record them." (NTSB, 1997, p. 93).

At 1806:14, the PF requested clarification "that's shaker and the stall?" and the flight engineer replied in the affirmative as "yeah, shaker and stall both." From that point, the FDR recorded that the airplane started to slow at a rate of about 1 knot per second toward the stall. The PF action to slow the airplane completed the decision loop from the flight engineer initiating the test point to the pilots' silent approval to proceed with the stall series (NTSB, 1997, p. 93).

5.2.4 Why Did it Make Sense for the ABX Crew to Proceed with the Stall Series

In hindsight, knowing the consequences of their decision, it would be easy to pinpoint that the ABX flight crew should not have attempted the stall series manoeuvres on 22 December 1996. As highlighted in the NTSB (1997) report, that fateful decision triggered a cascading series of events, which then led to the airplane impacting terrain in little over 3 minutes after the pilots decided to proceed with the stall-warning and stick-shaker functional test. Knowing what we know now, if the crew were to remove the stall series from the second evaluation flight, the same cascading pattern could have been prevented. However, framing the safety problem in a way that relies on knowing the accident outcome, is a counter-productive proposition in terms of mitigating the risk of performing similar functional tests on a DC-8 or a comparable design.

A more challenging (and more productive) question is to explore the reasons why it made sense for the ABX crew to proceed with the test point. This section of the case study revisits that critical decision, evaluating the decision context from the crew’s perspective. Figure 32 offers a graphical summary of systemic factors leading to the transition point.

Figure 32

Pre-transition Concept map for 1996 Narrows (continued overleaf)

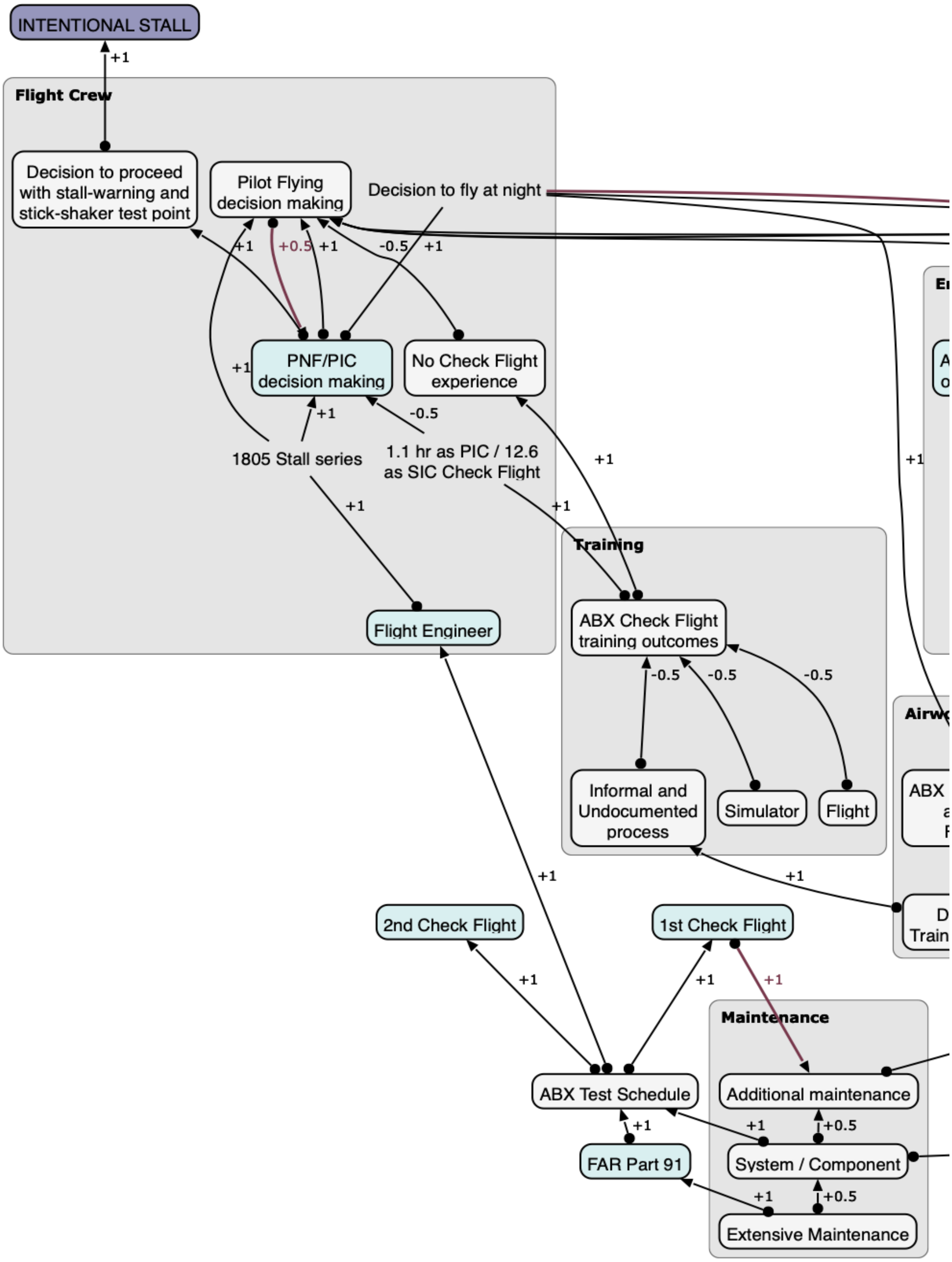
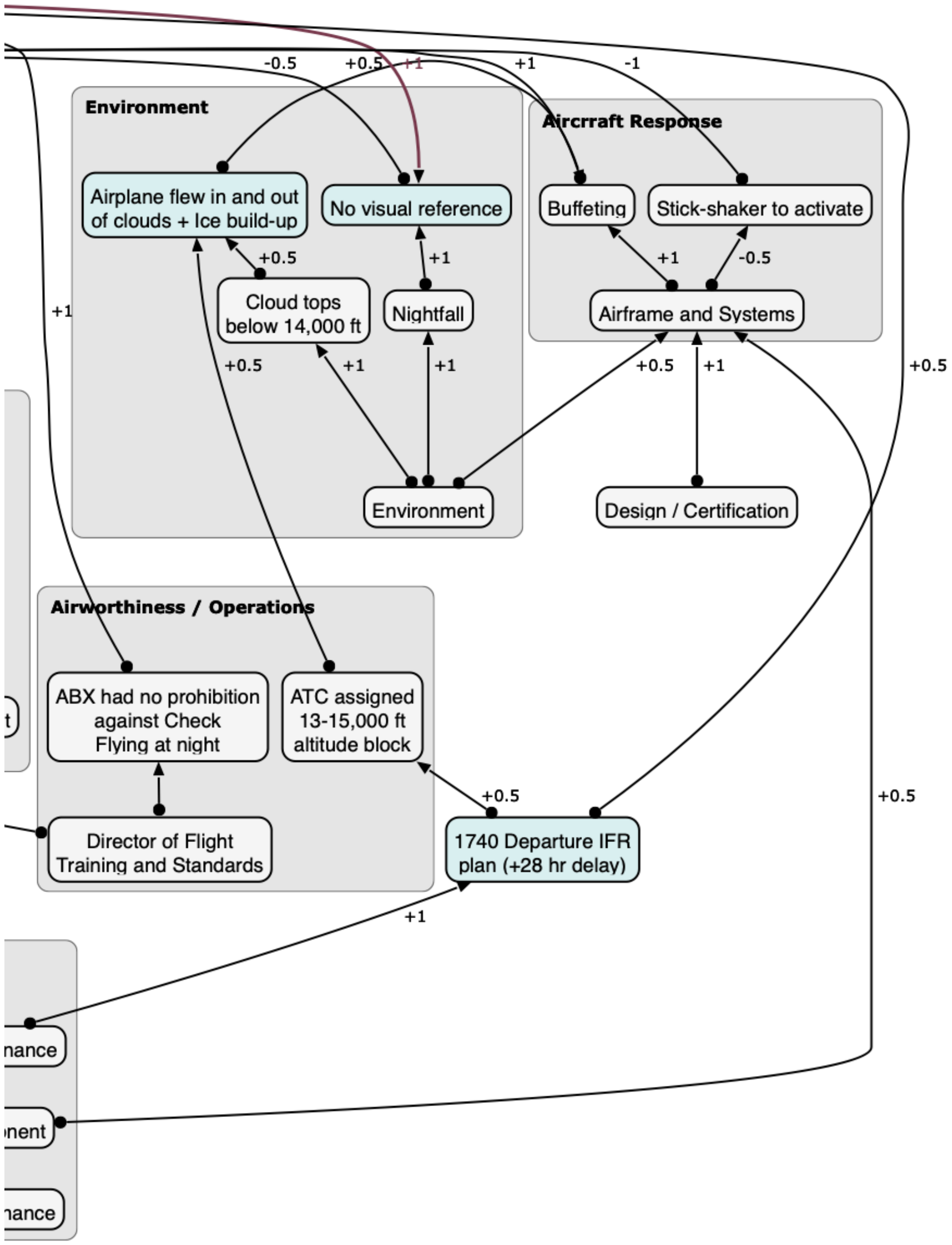


Figure 32

Pre-transition Concept map for 1996 Narrows (cont.)



Crew Experience and Composition. The PIC was an experienced training and check captain, and FAA-appointed DC-8 examiner, employed in a management role at ABX. The pilot’s management appointment made him responsible for commanding post-maintenance and post-modification evaluation flights, but he had very little direct check flight experience on the DC-8 (1.1 hour as PIC on the day before and 12.6 hour as non-flying second-in-command, previously). He had flown 15 check flights in DC-9 as PIC, some of which may have included an approach to stall (NTSB, 1997, pp. 4-6).

The PF was promoted to DC-8 flight standards manager in June 1996, replacing the PM in that position. Previously he was the DC-8 equipment chief pilot and was selected by ABX to become an FAA examiner. The PF had no DC-8 check flying experience prior to the partially completed test conducted on the day before the accident (NTSB, 1997, pp. 5-6).

The flight engineer was hired by ABX in 1988 and was promoted to DC-8 flight standards in 1991. He was also an FAA-designated examiner. He retired from the Air Force in 1986, after 26 years of service, which included multiple type ratings and instructor roles. Other than the day before the accident, his check flying experience was not disclosed in available records (NTSB, 1997, pp. 5-6).

The NTSB (1997, pp. 34-35) investigation evaluated CVR comments from the accident flight and technical logs from the previous check flight and determined that the PM was instructing the PF from the right seat and was serving as pilot-in-command (PIC) during both flights. All available evidence suggests that the PF was acting as second-in-command, therefore this study came to a similar conclusion about the unusual crew composition. Following on from that conclusion, the focus question is even more complex: why did it make sense for the ABX crew to proceed with the stall series during a combined check and training flight?

ABX Organisational Factors. The short answer to that question is that this is what they had always done within the ABX organisation. In his post-accident witness statement, the Director of Flight Training and Standards confirmed that ABX non-routine flight operations relied on an informal and undocumented system. The NTSB (1997, pp. 17-18) found that there was no formal crew training or selection program for pilots assigned to conducting check flight operations.

Both accident pilots were management pilots. In the ABX organisational and management framework they were responsible for their respective roles and expected to perform to the best of their self-assessed abilities. The PIC received his own DC-8 check flight instruction in a similar fashion, but in a non-flying capacity. His past DC-9 approach to stall experience, combined with the docile DC-8 stall response in the simulator, were supportive of the PIC's decision to proceed with the stall series. From an organisational view, the fact that two management pilots were expected to conduct a combined check and training flight – testing the automated stall warning and stick shaker functions and teaching the PM how to test those automated systems at the same time – did not make sense. This finding echoes the NTSB (1997, p. 51) causal determination about the ABX organisation's failure to establish and maintain an adequate formal program for check flight operations.

Environmental Factors. For an external observer, the decision to proceed with the stall series in the actual environmental conditions experienced on that fateful day is a questionable one. The NTSB explored the crew's earlier decisions to fly at night with no visible horizon and the decision to continue despite witnessing ice build-up on the airframe. The NTSB (1997, pp. 48-49) concluded that the lack of a visible horizon was a factor in the accident, while the potential role of airframe icing was discounted by a retrospective analysis of expected and actual stall speeds recorded.

From the crew's perspective, the decision to conduct the test after nightfall needs to be put into an organisational and operational context. While airframe manufacturers did recommend not to perform this type of in-flight validation without a natural horizon, the ABX crew was most likely not aware of that fact, as supported by post-incident interviews. They operated in an organisational bubble, where there were no formal requirements to program the check flight so that all required test points are completed before end of local daylight (NTSB, 1997, pp. 42-43).

On the other hand, the CVR recording provides evidence that the PF was concerned about ice progressively building up on their airframe when they were flying in and out of clouds in the assigned airspace block. While the NTSB's (1997, p. 48) engineering analysis was correct that the early buffeting experienced during the approach to stall sequence may as well had been caused by incorrect flight control rigging, it did not make sense for an experienced crew to proceed with the stall series when ice build-up was present.

The PF repeatedly expressed his concerns, however there is no evidence that his mental model was shared with the rest of the crew. To the contrary, the fact that the PF did not object to slowing down the aircraft when approaching the stall indicates that he either did not want to challenge the decision to proceed, or he was satisfied that the risk was acceptable. The PF did not speak up when the PF expressed his surprise that the airframe was buffeting much earlier than expected. For an experienced test crew, this is a textbook scenario of applying "knock-it-off" criteria and calling for a stop in the test program. This is not to say that the PF should be blamed for not making the call, rather to highlight that this crucial decision making aspect was not explored in the NTSB (1997) report.

Applying a set of knock-it-off criteria during flight tests or check flights is a common and effective risk control measure, when agreed to and understood by all crew members during the pre-flight briefing, at the latest. Neither the NTSB (1997) report, nor the adopted safety recommendations considered this fundamental risk management framework, which appears to indicate that the DC-8 accident was evaluated in an airline operation context, not as a post-maintenance test flight.

Stall Warning System. Another critical aspect in the crew's decision to proceed with the stall series is the correct functioning of the automated stall warning and stick shaker subsystems. The post-accident NTSB (1997, pp. 36-37) engineering analysis highlighted several valid concerns about the DC-8 stall warning system's reliability, maintainability and overall system safety, including the early generation warning technology, the designer's decision to calibrate the system in a "set and forget" style (the last full calibration for the accident aircraft was recorded in 1967 at the factory), and the limited ground tests required when the system was disturbed during maintenance.

The NTSB (1997, p. 8) also found that the pre-flight checks of the stall warning system were very limited in scope. The DC-8 design is based on a lift transducer, located in a leading edge of the wing, as the sensing mechanism of the stall warning function. A stall warning test is performed during pre-flight, however that only tests the electrical integrity of the lift computer and stick shaker activation. The lift transducer is not tested during cockpit checks. Similarly, the only flight deck effect for a failed lift transducer heater is a “Heater INOP” light, when the current measured in the heater circuit is outside a set tolerance window.

The poor reliability of the early generation DC-8 stall warning and stick shaker systems was common (tribal) knowledge amongst airline pilots and the accident crew was most likely no exception. Unless brought up by maintenance engineers during a pre-flight briefing, none of the other engineering issues identified by the NTSB would concern the flight crew, though. The characteristic small airspeed margin between stick shaker activation and the stall, combined with the programmed behaviour experienced during simulator sessions, would have reinforced a bias to continue approaching the stall, even if the stick shaker did not activate.

Decision to Fly at Lowest Assigned Altitude. The Safety Board observed that the crew decided to conduct the stall series only 500 ft above the bottom of the assigned altitude block on top of clouds. The investigation also found that they obtained clearance for a 2,000 ft block, instead of a 3-5,000 ft block preferred for stall recoveries. In hindsight, knowing the critical importance of trading available altitude for more time during an upset recovery attempt, the NTSB (1997, pp. 2, 106) comments make perfect sense.

From the accident crew's perspective however, there is no evidence to suggest that they were trained to follow a modified stall recovery procedure which targeted a rapid stall recovery in lieu of the traditional focus on minimum altitude loss. In 1991, after a similar LOC-I incident during a DC-8 check flight, the FAA directed ABX to introduce the modified stall recovery technique (NTSB, 1997, pp. 105-106). The then-flight standards manager adopted the new technique, however when he returned to the line in 1994, he was replaced by the accident PM, who was trained by the flight technical director using the old recovery technique. In other words, the informal ABX organisation was causal in losing a flight safety critical risk control, not the accident pilots.

5.2.5 Significant Events During the Transition to Recovery

While the airplane was slowing down, the main objective of the flight again reverted to in-flight training from the original goal of testing the stall series. At 1806:18 the PM initiated the exchange when he highlighted that the engine rpm should not slow to flight idle during the manoeuvre, telling the PF that "the only trick to this is just don't unspool." (NTSB, 1997, p. 93).

At 1807:21, the PF requested further clarification about correct engine power, when he asked, “are you saying you don’t want to pull all the way back to it [the stall] and then spool back or just wait?”. In response, the PM stated, “Aw you can do that, just when you get close to the stall you don’t want to be unspooled.” The PF then requested help from the PM in setting correct engine power when he replied, “Unspool and then I’ll respool.” (NTSB, 1997, p. 94).

Shortly after that exchange, the PF alluded to his preferred course of action by saying “Yeah, I’m going to spool now.”. The PM confirmed that course of action by a short “All right” (NTSB, 1997, p. 94). According to the NTSB (1997, p. 95), the CVR recorded sounds similar to the engines increasing in rpm at 1807:55.

At 1808:06, the airplane was flying at 151 knots when the PF announced, “some buffet” and the PM expressed some surprise by noting “yeah, that’s pretty early.” (NTSB, 1997, p. 95).

At the 1808:09 timestamp, the CVR recorded the sound of rattling.

At 1808:11, the airplane was flying at 145 knots when the flight engineer said, “that’s a stall right there… ain’t no shaker” (NTSB, 1997, p. 95).

5.2.6 Significant Events During the Recovery Attempt

At 1808:13, the PF called “set max power” which reflects a timely decision to start recovering from the stall. At 1808:20 the CVR started to record popping sounds which continued for nine seconds (NTSB, 1997, p. 95).

At 1808:30, the PM gave direct – but ambiguous – instructions to the PF, when he stated “You can take a little altitude down. Take it down”, then adding at 1808:42, “Start bringing the nose back up.” (NTSB, 1997, p. 96).

At 1809:10, based on primary radar returns, ATC contacted the ABX crew, querying whether they were in an emergency descent. The PM replied with a short “yes sir” in the affirmative. ATC then followed up with “can you hold seven thousand?” to which there was no reply. ATC received no further radio calls from the ABX crew (NTSB, 1997, p. 98).

At 1809:29, the PM intervened again, telling the PF to apply left rudder. At 1809:30 the PF replied, “left rudder’s buried”. At 1809:32, the PM added “OK, easy, don’t. OK now, easy bring it back.” (NTSB, 1997, p. 99). At 1809:35 the GPWS activated and the aural warning “terrain, terrain, whoop, whoop, pull up” was recorded on the CVR. Three seconds after the GPWS warning, or 1 minute and 32 seconds after the PF observed some early buffeting, the airplane impacted mountainous terrain in a 26-degree nose-down and left-wing low attitude at about 3,400 feet AMSL (NTSB, 1997, p. 13).

5.2.7 What Systemic Factors Contributed to the Failed Recovery Attempt

The focus question in the second part of the case study is more aligned with the causal analysis performed by the NTSB (1997). The main difference, as illustrated in Figure 33, is the addition of negative causal links, which guide the analyst’s attention to causal factors which did not support the crew’s decision making process or indicate interactions which may have triggered unexpected cascading failures propagating in the system.

The NTSB (1997, p. 51) nominated the inappropriate control inputs applied by the PF and the failure of the PM to recognise, address, and correct the inappropriate recovery technique during the stall recovery attempt as immediate causes of the accident. Furthermore, the Board also nominated the inoperative stick shaker and the inadequate fidelity of the ABX DC-8 flight simulator as contributing to the failed recovery.

Figure 33

Post-transition Concept map for 1996 Narrows (continued overleaf)

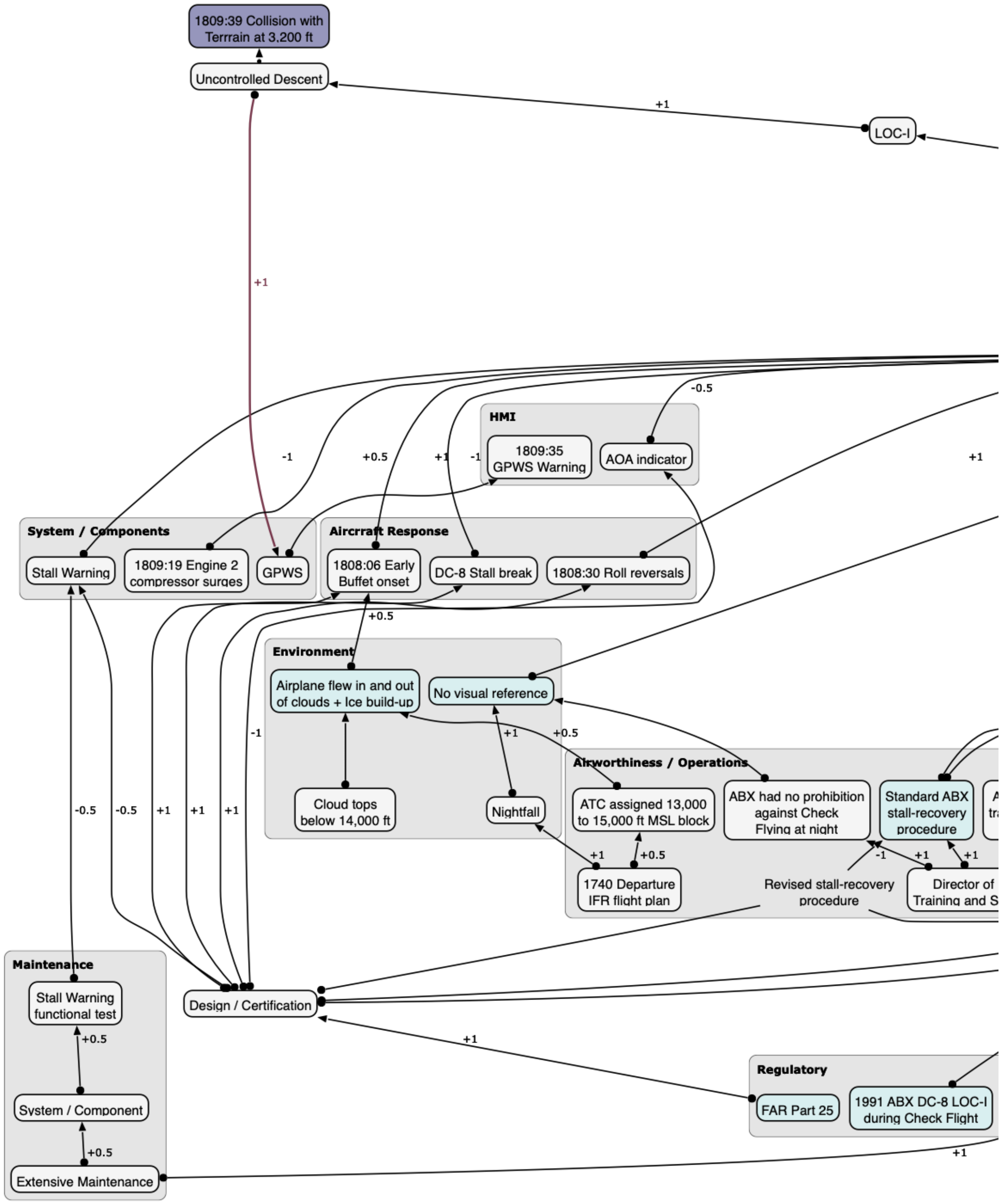
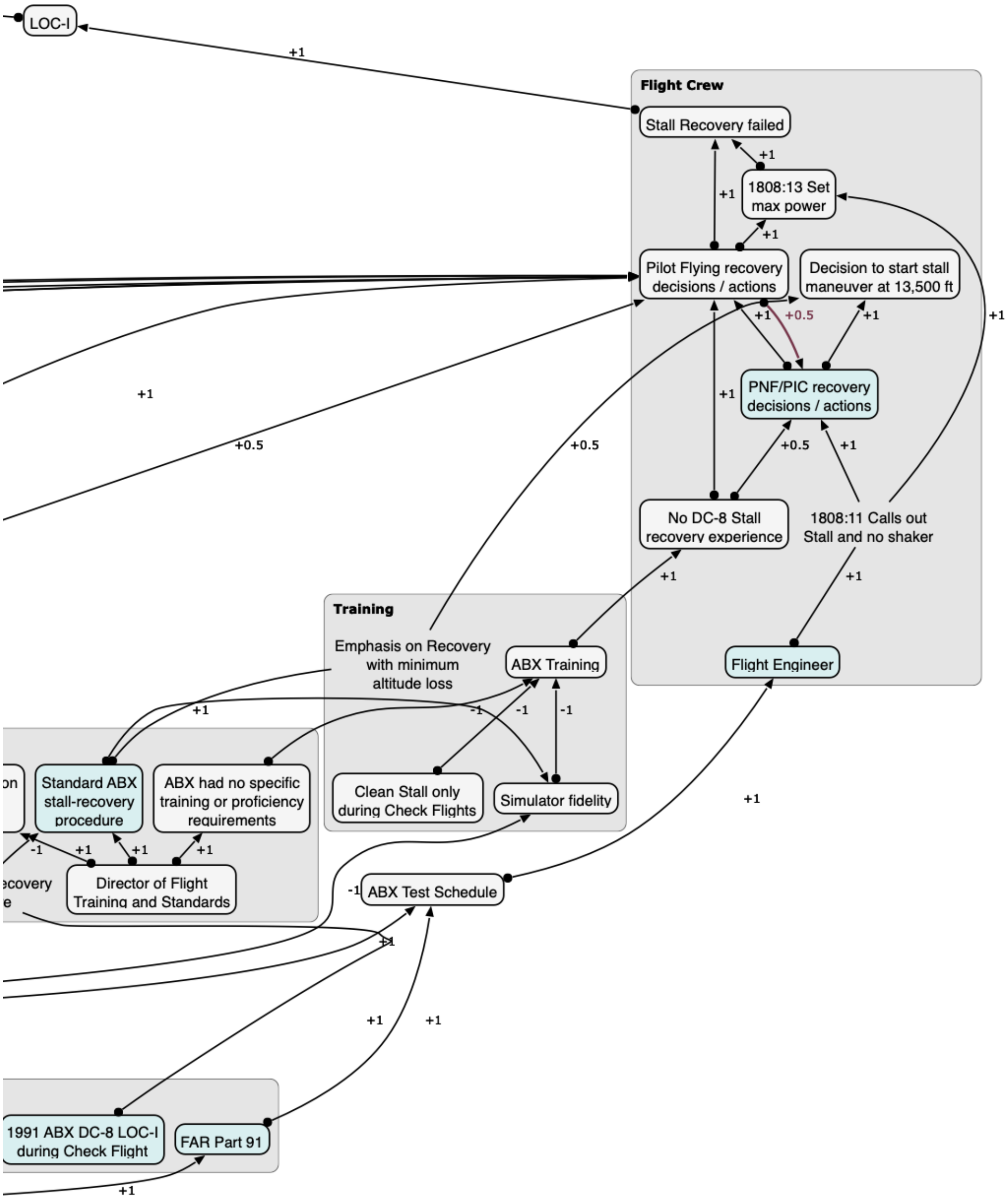


Figure 33

Post-transition Concept map for 1996 Narrows (cont.)



There is no intent to revisit the causal factors annotated by the NTSB, including the probable contribution by the failure of the ABX organisation and flight operations management team. This analysis will focus on the differences between the NTSB causal analysis and the causal patterns uncovered when accounting for negative causal relations. Working backwards from the loss of control, the systemic causes in the [Flight crew], [Training], [Airworthiness – Operations], and [Environmental] causal groups were explored in the previous section and will not be repeated here.

The Safety Board's engineering analysis did mention several concerns about the DC-8 airframe and flight deck design, as well as power-plant integration, however none of those concerns were annotated as probable causes or factors contributing to the accident sequence (NTSB, 1997, pp. 37-39). This is an unfortunate limitation of the narrow RCA technique imposed on the Board by its legislated mandate.

Contrary to the NTSB (1997) report, this analysis found that the type-certified DC-8 design did introduce systemic causal factors which propagated from the design and certification phase, all the way to in-service operations, and finally the accident flight. Prime examples are the early generation stall warning system, the lack of AOA feedback to the pilots, the compressor surges and sensitive power management required during a stall, and the simulator certification shortfalls in reproducing clean stall behaviour and roll reversals. Furthermore, while generic FAR Part 91 rules required the operator to perform a functional test after extensive maintenance, which ABX complied with in this instance, the regulatory framework was not robust enough to mitigate the propagating effect of potential critical maintenance errors, such as the notoriously unreliable DC-8 stall warning and stick shaker system.

5.3 Within Case Analysis

5.3.1 Introduction

Following on from reviewing the categories, concepts, and causal relationships in the pentavalent causal maps, this part of the case study elicits basic patterns observed in the data. As explained in chapter three, the intent is to review the three basic patterns in how adaptive systems fail, namely decompensation, locally adaptive and globally maladaptive responses, and adaptive units getting stuck in outdated behaviours (Woods & Branlat, 2017).

To achieve that aim without a complex network model of adaptive units, the case study assumes that a simple interaction model can adequately describe the main controllers, including a) the crew, b) the automated systems, c) the ABX Flight Standards unit, d) the ATC unit, e) the Maintenance Repair and Overhaul (MRO) organisation, f) the airframe manufacturer, and g) the FAA oversight unit. The list is in order from lower to higher level adaptive units. The crew is the only adaptive unit where a refinement was necessary to differentiate between the PF and PM.

Table 4 provides a high-level summary of the key events and basic (mal)adaptive patterns revealed from the pentavalent causal maps. The table introduces time as a critical dimension for failure analysis. Following the modified resilience terms introduced by Woods (2018), the analysis follows key challenges and surprises (events), the actions deployed (response), the risk of saturation (compensation), any potential misalignment or mismatch between the units (coordination), and any evidence of being stuck in outdated models when the environment has changed, or a new disturbance was introduced (revised model).

Table 4

Basic Patterns in Adaptive System Failures – Case A

Time	Key Event	Response	Compensation	Coordination	Revised model
Year 1967	Last full calibration	Manufacturer GO	Operations	Potential mismatch	Outdated model
Year 1991	LOC-I event during check flight	FAA direction	Modified stall recovery		Updated training
Year 1994	New FLT STDs Mgr	Stall training	Old recovery technique	Mismatch	Outdated training
Pre-flight	Limited stall tests	Crew decision GO	Silent transfer	Mismatch	Outdated model
1740K	Late departure	Crew decision GO	No visual reference	Mismatch	Outdated plan
1742K	ATC assigned block 13-15,000ft	Crew decision GO	Min margins & 14,000 cloud tops	Mismatch	Outdated plan
1748:34K	Ice buildup	Crew decision GO	Stay out of clouds	Potential mismatch	Outdated plan
1806:14K	Stall series	Crew decision GO	500 ft above clouds Lowest altitude	Mismatch	Outdated model
1806:18K	In-flight training	Crew exchange	PM to set PWR	Potential mismatch	Outdated plan
1808:06K	Early buffet	Crew decision GO	Control transfer	Mismatch	Outdated model
1808:11K	Stall at 145 kt	Crew decision GO	Silent transfer	Mismatch	Outdated model
1808:13K	Recognition	Pilot input / Thrust	Set MAX power		Updated model
1808:42K	Verbal instruction from PM to PF	Pilot input / Elevator / Nose up	Saturated	Cross-purposes	Outdated model
1809:10K	Emergency descent	ATC query	Saturated	Cross-purposes	Outdated model
1809:29K	Apply left rudder!	Pilot input / Rudder	Saturated		Outdated model
1809:35K	GPWS warnings	Automation	Decompensation		

5.3.2 Adaptive System Failures

Design and calibration – the accident airplane was manufactured in 1967. That was the last time the early generation stall warning system was fully calibrated. Airline flight and maintenance operations compensated for the manufacturer's decision, however the potential for surprise remained throughout its service life. The poor reliability of the DC-8 stall warning system was tribal knowledge, accepted as a design characteristic by pilots and maintenance engineers.

Organisational failure – in 1994, the ABX Flight Standards manager was replaced by the pilot who commanded the accident flight (PIC). The PIC received an outdated DC-8 stall recovery training that contradicted and invalidated FAA directions received three years earlier, in response to a similar LOC-I incident at ABX. The outdated recovery technique, aggravated by the DC-8 flight simulator's flawed representation of the actual stall response, was responsible for the PIC's inability to properly monitor and intervene prior to the airplane entering a deep stall.

Silent control transfer – the limited pre-flight stall warning system checks did not provide an opportunity for the crew to discover that the system was unserviceable. Unknown to the crew, the automated system transferred the responsibility to them for monitoring stall margins by other available means (speed margins, buffeting, airframe noise, visual reference). That mismatch became critical during the approach to stall phase.

Safety margins sacrificed – the crew made a series of decisions (late departure, lowest acceptable airspace block, ice build-up) that eroded critical safety margins, such as a natural horizon, a safe altitude, and comfortable stall margins (compensating). In hindsight, any one of those decisions would have justified not entering the stall series test point in their program (outdated plan).

No capacity to adapt – the decision to proceed with the stall series marked the point where the accident flight started shifting from a check flight to a recovery phase. The crew had no margin for error when approaching the stall, and at the same time, they were facing a significant challenge event in the form of a LOC-I scenario that they have never experienced in a DC-8.

Communication and coordination breakdowns – during the deceleration phase, the PF wanted to delegate thrust management to the PIC, repeatedly requesting clarification about the correct recovery technique. Shortly after, the airframe signalled the approaching stall by an early onset of buffeting that was recognised by the crew. The PIC did not take control and did not knock-off the stall series test. It remains plausible that the crew was expecting stick shaker activation and a benign stall response, a mental model shaped by their simulator experience.

Recognition – at the point of initial stall entry (145 kt), the airplane decompensated and switched to the crew for preventing deep stall entry. The PF recognised the initial stall entry and initiated recovery by applying maximum thrust. Considering that this was his first full stall entry with the DC-8, acting as a trainee check pilot, the PF must be given credit for the recovery attempt. There was no exchange between the pilots about the stall recovery technique prior to experiencing the deep stall and roll reversals.

Saturation and a new surprise – for the remaining part of the flight, the crew was at (or close to) saturation. It is important to note that the crew was working at cross-purposes with the airplane when repeatedly applying nose up input to the elevators. The PIC gave priority to an ATC transmission while his trainee could not regain control of the airplane. The fact that the rudder was completely ineffective (buried) created a fundamental surprise for both pilots.

Time – the crew was under considerable time pressure during the stall series test. From the moment of slowing down the aircraft to the PF initiating the stall recovery, less than 2 minutes passed. There were an additional 82 seconds from initiating the stall recovery to the fateful GPWS terrain warning. It was simply not possible for the crew to keep up with the pace and tempo of events in a rapidly deteriorating scenario (decompensation).

5.4 Discussion of the Findings

Focus Question One: Why did it make sense for the crew to continue the test program?

In summary, while the crew members were very experienced DC-8 operators, most of their hours were logged during routine line operations. A series of adaptive traps prevented them from realizing that it was ill-advised to proceed with the stall series. Unknown to the pilots, the early generation stall warning system was not calibrated for almost 30 years, and it was unserviceable during the accident flight. The limited pre-flight checks did not provide an opportunity for them to realize that the system was unserviceable. The ABX organisation did not provide them with effective risk controls prior to the accident flight. The decision to continue with the stall series at night, with a known ice build-up on the airframe and at the lowest acceptable altitude bracket (only 500 ft above clouds), indicates that the crew was not expecting any major complications or considerable altitude loss.

Focus Question Two: Why did the recovery attempt work / fail to work?

The PF and PIC/PM were experts in routine flying, but not in a check flight decision making context. They were attempting a complex test point for which neither pilot was adequately trained. The PIC's DC-9 check flying was noted as relevant experience, however there is no record of the PIC ever commanding a full approach-to-stall and recovery sequence on a DC-8. The accident flight served as a training opportunity for the PF, a secondary mission objective that was not fully aligned with the main goal of safely returning the airplane to service after a major modification program. Once the crew made the decision to enter the stall series at night, with no visual horizon and at a relatively low altitude, they sacrificed their remaining safety margins. From the point of stall entry, the crew was saturated and working at cross-purposes with the airplane. They simply could not cope with additional surprises, like the airplane's unexpected stall response or the ineffective (buried) rudder controls. Their inability to recover from the stall reflects systemic failures by the regulator, the manufacturer, and the airline operator.

The case study revealed basic adaptive system failures and patterns when the primary control units interacted with each other, including the a) the crew, b) the automated systems, c) the ABX Flight Standards unit, d) the ATC unit, e) the MRO, f) the airframe manufacturer, and g) the FAA oversight unit. The unreliable stall warning system design, coupled with local adaptations during design and operations, resulted in a brittle system with little or no capacity to adapt to new challenges. The problems manifested in coordination breakdowns, including sudden and silent control transfers.

Ultimately, in the context of check flying, the case illustrates a fundamental problem with the inherent design assumption that automation can always fall back on the human pilot as a last resort. When the pilot is already struggling to keep up with the pace of events in a rapidly changing environment, it is highly unlikely that the additional demand can be met in the middle of a challenge, unless new forms of system and coordination failures are mitigated by design and operational risk controls.

Chapter 6 – CASE B: A320 AOA PROTECTIONS

6.1 Introduction

Event: Accident – Airbus A320 registered D-AXLA, operated by XL Airways Germany

Location: Off the coast of Canet-Plage (on approach to Perpignan)

Date: 27 November 2008

Agency: BEA

On 27 November 2008, an A320 operated by XL Airways Germany (GXL) was destroyed when the airplane impacted the sea near the coastline in Southern France. There were three pilots, three maintenance engineers, and an airworthiness inspector on board, all of whom lost their lives upon impact. The end of lease functional test flight (check flight) departed from Perpignan. After about an hour of flight, the airplane returned to Perpignan airspace, and was then cleared to carry out an ILS approach, followed by a go around and a departure towards Frankfurt. Shortly before overflying the initial approach fix, the crew carried out the test point for AOA protections in normal law. They lost control of the airplane, which crashed into the sea (BEA, 2010).

The final report concluded that the following causes and contributing factors led to the accident:

> The accident was caused by the loss of control of the airplane by the crew following the improvised demonstration of the functioning of the angle of attack protections, while the blockage of the angle of attack sensors made it impossible for these protections to trigger. The crew was not aware of the blockage of the angle of attack sensors. They did not take into account the speeds mentioned in the programme of checks available to them and consequently did not stop the demonstration before the stall (BEA, 2010, p. 98).

According to the BEA (2010, p. 98) investigation, several factors contributed to the accident: a) the crew's decision to carry out the demonstration at a low height; b) the PF incorrect recovery technique and failure to manually trim the airplane; c) the ad-hoc management and improvisation during the test program; d) the decision to use a schedule developed for flight test purposes; e) the absence of a regulatory framework for non-revenue flights; and the ambiguous AMM instructions for rinsing the airplane that led to water entering the body of AOA sensors 1 and 2. Finally, the investigation also found that the following factors likely also contributed to the accident: f) inadequate coordination between an atypical team composed of three airline pilots in the cockpit; and g) fatigue that may have reduced the crew's awareness of actual system states (BEA 2010, p. 98).

In an earlier phase of the research study, a text analysis was applied to the BEA (2010) investigation report that summarized the logical and temporal structure of causal and contributing factors (see Figure 34). The next step of the research project built high level concept maps, based on the main causal groups and causal links extracted from the BEA report. When the higher-level abstraction in Figure 22 is compared to the text analysis results in Figure 34, it is obvious that causal details are less refined, as the abstraction hides the direct edges between individual causal factors. However, when the two maps are compared, a linear translation can reveal those hidden details.

For example, the BEA's (2010, p. 98) causal statement about the "absence of a regulatory framework in relation to non-revenue flights in the areas of air traffic management, of operations and of operational aspects" translates to the causal link between [MCF Regulations] and [Airworthiness/Operations] nodes, and the +0.5-causal strength (contributing causal factor) assigned to the edge between those nodes.

Figure 34

Text Analysis for 2008 Perpignan

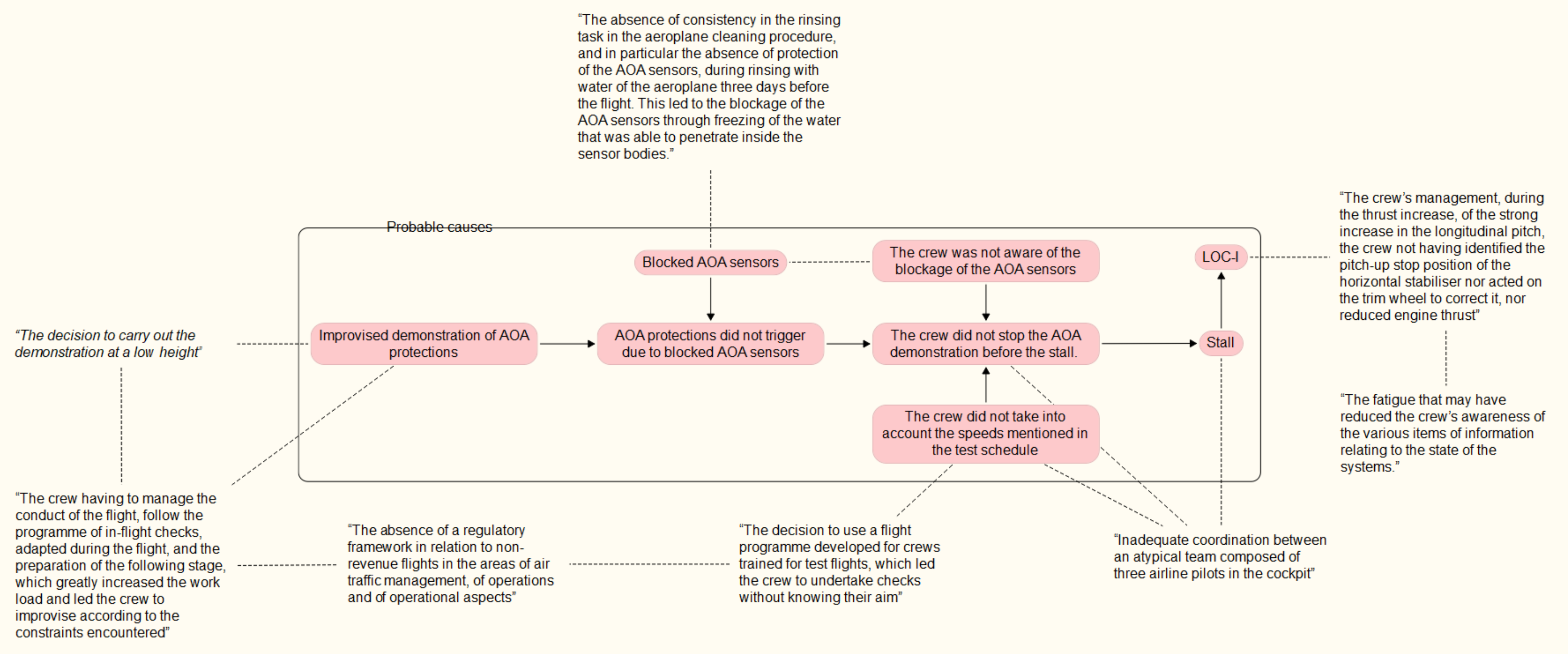
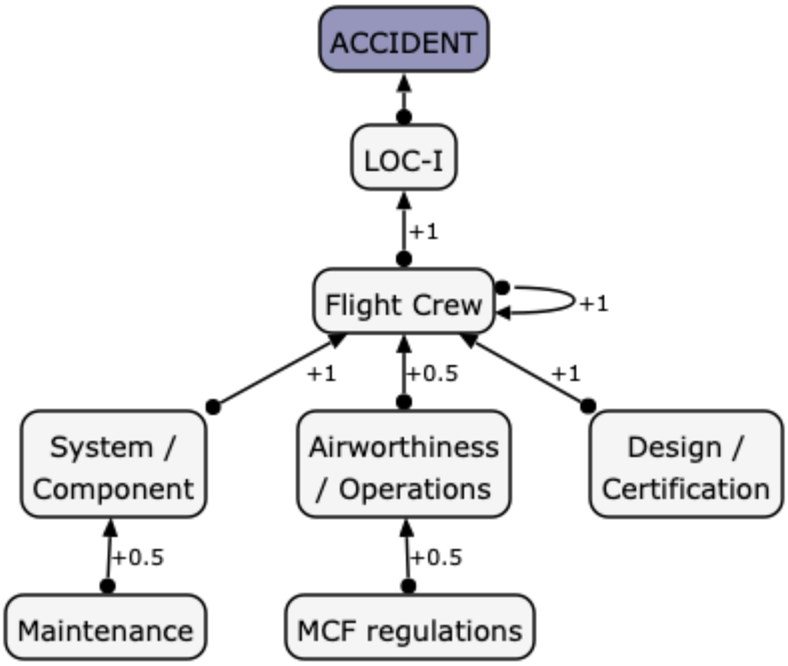


Figure 22

A04 – 2008 Perpignan (duplicated for ease of reference)



Similar to the French Bureau's mission, the sole objective of this case study is to draw lessons from the Perpignan accident. As highlighted in the literature review chapter, unlike the NTSB, BEA investigations do not seek to identify probable causes. The French Bureau's final reports generally conclude by annotating findings as either causes, contributory factors, or probable contributions to the accident sequence.

The case study relies on the final report issued by BEA (2010); however, it has a slightly different scope, guided by two focus questions. In this instance, the case study intends to reveal a) the reasons why it made sense for the accident crew to proceed with testing low speed flight envelope protections at a relatively low height and b) the reasons why the recovery attempt was not successful.

6.2 Data Analysis

6.2.1 Nominated Transition Point

Based on the sequence of events described in the Bureau's final report, the decision to proceed with testing the low-speed flight envelope protections at a low height was nominated as the point where the check flight started to transition from the original test program to a recovery phase (BEA, 2010, p. 86). The case study describes noteworthy events leading up to the accident flight and significant events pre- and post-transition. The pre-transition narrative focuses on factors contributing to the crew's decision to accept and start the low-speed test, inadvertently stalling the airplane at a low height. The post-transition narrative revisits available evidence in the context of their failed recovery attempt, including the point when the crew realized the need for recovery and initiated the attempt.

6.2.2 Noteworthy Events Prior to the Accident Flight

On 3 November 2008, GXL had ferried the accident airplane to Perpignan for a base maintenance check and external painting. The certificate for release to service was issued by the Part 145 MRO on 27 November 2008. The airplane was at the end of a lease agreement, returning to Air New Zealand (ANZ). As common in the industry, the lease agreement specified a series of in-flight functional checks prior to the owner accepting the airplane's return to its fleet and line operations (BEA, 2010, p. 15).

The crew complement was a mix of GXL and ANZ personnel, including a GXL captain (PF) and copilot (PM), and an ANZ check flight captain acting as an observer on the flight deck, three maintenance engineers from ANZ, and an airworthiness inspector of the New Zealand Civil Aviation Authority. Prior to take-off, the ANZ captain confirmed his observer role to the German pilots by requesting them to raise a hand, should he interrupt their flying duties during the test program (BEA, 2010, p. 15).

The original departure time, according to flight plan records, was 1230Z for a planned 2h 35m duration in the South-Western (SW) sector of France, returning to Perpignan controlled aerodrome airspace, only to continue to Frankfurt in Germany for formal handover to ANZ. The actual departure was delayed to 1433Z due to last-minute maintenance actions (BEA, 2010, p. 15).

Five hours earlier, a GXL B737, with a flight plan that was identical to the accident aircraft, had contacted ATC on several occasions to be able to perform check flight manoeuvres that had required extensive coordination between the different control sectors (BEA, 2010, p. 15).

6.2.3 Significant Events Leading up to the Transition Point

The aircraft got airborne at 1444Z. Minutes later, the ATC controller for the SW sector contacted Perpignan approach via phone to ensure that the accident flight had the relevant authorizations for a plan he referred to as “a disguised test flight”, like the B737 check flight earlier in the day. When the GXL crew contacted the SW sector to conduct a 360 turn, the ATC controller rejected their request, stating that it was not compatible with standard rules of flight in general controlled airspace. In response, the GXL crew confirmed that they wanted to continue as per the planned route (BEA 2010, p. 15).

At CVR timestamp 1454:25, the PF suggested to the ANZ observer to delay flight control checks in normal law until the approach segment, when returning to Perpignan airspace. The proposal was not objected to by the ANZ observer. The checks in normal law were performed by the PF at around 1512Z on an opportunity basis (BEA 2010, p. 16).

At around FDR timestamp 1503Z, the airplane was flying at FL320. Shortly after reaching their cleared altitude, local AOA values from AOA sensor 1 and 2 stopped at almost identical values which corresponded to the cruise AOA of the airplane. The crew did not have direct feedback about the incorrect AOA values, which did not change until the end of the check flight (BEA 2010, p. 16).

At CVR timestamp 1522Z, the flight was cleared to FL390 for a planned high altitude Auxiliary Power Unit (APU) start. While waiting for the clearance, a conversation ensued between the ANZ pilot and the GXL captain regarding the few remaining test points for the descent phase, the option to delay the APU start until the flight was on its way to Frankfurt. The ANZ pilot recited the low speed test point, stating that the test point should be carried out at FL140, but failed to mention the critical airspeed values involved in the test (BEA, 2010, p. 16).

At 1533:22, having been cleared to descend to FL130, the ANZ pilot requested the flight control test in alternate law, scheduled in the test program above FL140. The PF decided to carry out the test at FL130, which was accepted by the ANZ pilot. At 1533:34 the flight was cleared to descend further to FL120 and had been requested to slow down to 250 kts, then to 200 kts, and plan for holding at the Perpignan VOR (BEA, 2010, p. 16).

At CVR timestamp 1536:47, the flight was level at FL120, the PF asked, “you want alternate law” and the ANZ pilot responded in the affirmative. The alternate law test was performed at FL120 (in lieu of the FL140 minimum prescribed in the test manual) prior to descending further to FL080 and speed to 180 kts (BEA 2010, p. 16).

At 1537:52, the flight control laws returned to normal law, indicating the end of alternate law test, and autopilot 1 was reconnected.

The ANZ pilot suggested the low-speed test point as the next item, then described the sequence of events for the check at low speed in full configuration. The PF asked if the test intent was to go down to VLS and alpha floor protection. The ANZ pilot responded in the affirmative and requested the PF to pull quite hard when reaching VLS to activate the alpha floor protection. The PF confirmed that he was fully aware. Then the ANZ pilot added additional clarification when he said "then you need to pitch forward and err… you're happy with disconnect and reengage. And out of alpha floor". Again, neither the minimum safe altitude, nor the limit speeds were discussed or even mentioned between the crew members (BEA 2010, p. 17).

At CVR timestamp 1538:52, the flight was cleared to descend to FL060. At that point, the aircraft was flying below FL 100 at 214 kts. At around 1540K, Perpignan approach requested a turn to heading 190 while maintaining 180 kts. About a minute later, ATC cleared the flight for the LANET- ILS approach to RWY 33 and to descend to 5,000 ft. At that point, the PF told the ANZ pilot that the low-speed check would either need to be delayed, or only performed during the flight to Frankfurt. The PF also raised the possibility of not performing the low-speed test point at all (BEA 2010, p. 17).

At 1542:14, Perpignan approach asked the crew to report their airspeed. The PM initially stated that it was falling, then a few seconds later reported 180 kts. At that point, the aircraft was flying slightly above 190 kts and the selected speed was 157 kts. Approach asked to maintain 180 kts and cleared the flight to 2,000 ft.

At 1542:46, the PF stated that the LANET approach was not available in the FMS database. The PM performed an approach brief 36 seconds later. Between 1543:20 and 1543:55, the spoilers were deployed by the crew (BEA 2010, p. 17).

At 1543:37, the PF disengaged the autopilot, then he turned to the ANZ pilot with "Down below the clouds, so you want what?". The ANZ pilot responded with questioning "to go slower you mean?". Both the PF and PM responded in the affirmative. At 1543:41, the PF reduced the thrust to flight idle and the auto-thrust disengaged. The ANZ pilot added with an ambiguous "We need to go slow with err recovery from… recovery". The airplane was flying at 4,080 ft and at 166 kts (BEA 2010, p. 17).

At CVR timestamp 1543:48, the PF requested the PM to extend the landing gear then said, "we do the err the…". The ANZ pilot completed the PF sentence by saying "low speed yeah" (BEA 2010, p. 18). At that moment, the decision loop between the GXL and ANZ captain about performing the low-speed test at a low height was completed, and the crew started the transition to the recovery point.

6.2.4 Why Did it Make Sense for the Crew to Proceed with the Low Speed Test at a Low Height

In retrospect, knowing what happened a few minutes later, it would be easy to state that the accident crew should not have attempted to perform the low-speed protections test during the approach phase to Perpignan, especially at such a low height. As annotated by the Bureau (BEA, 2010, p. 96), their decision propagated a series of events, which ultimately led to the airplane impacting the surface of the sea in just over 2 minutes after the PF decided to proceed with the low-speed protections test. The recovery attempt lasted 62 seconds from the first stall warning to impacting the surface.

Knowing full well the catastrophic consequences, of the crew were to defer the low-speed protections test to a later segment of the check flight, they may have had a better chance of recovering from the upset condition. The blocked AOA sensor 1 and 2 would have presented a similar major challenge to both automated flight systems and the crew, however had more altitude margin been available, they would have had a chance to trade altitude for more time. Framing the crew's decision as part of a causal chain introduces a considerable risk of hindsight bias on the investigator's behalf.

As highlighted in the previous case study, a more productive question (in a safety prevention context) is to explore the reasons why it made sense for the GXL captain to endorse proceeding with the test point. This section of the case study revisits that critical decision loop, evaluating the decision context from the crew’s perspective. Figure 35 offers a graphical summary of systemic factors leading to the transition point.

Crew Experience and Composition. The PF was the head of air and ground operations for XL Airways, an experienced A320 family captain, instructor, and type rating examiner. He had a very limited exposure to check flying when he conducted an acceptance flight in Toulouse for another operator in 2004. The PM was an experienced first officer at GXL, who partly completed a captain upgrade program under the supervision of the PF, but later ended the upgrade training due to personal reasons. He had no recorded check flying experience. Both GXL pilots were based at Frankfurt, and on the day of the accident flight they left home very early in the morning to position to Perpignan by taxi, a commercial flight, then rental car (BEA 2010, pp. 21-23).

Figure 35

Pre-transition Concept map for 2008 Perpignan (continued overleaf)

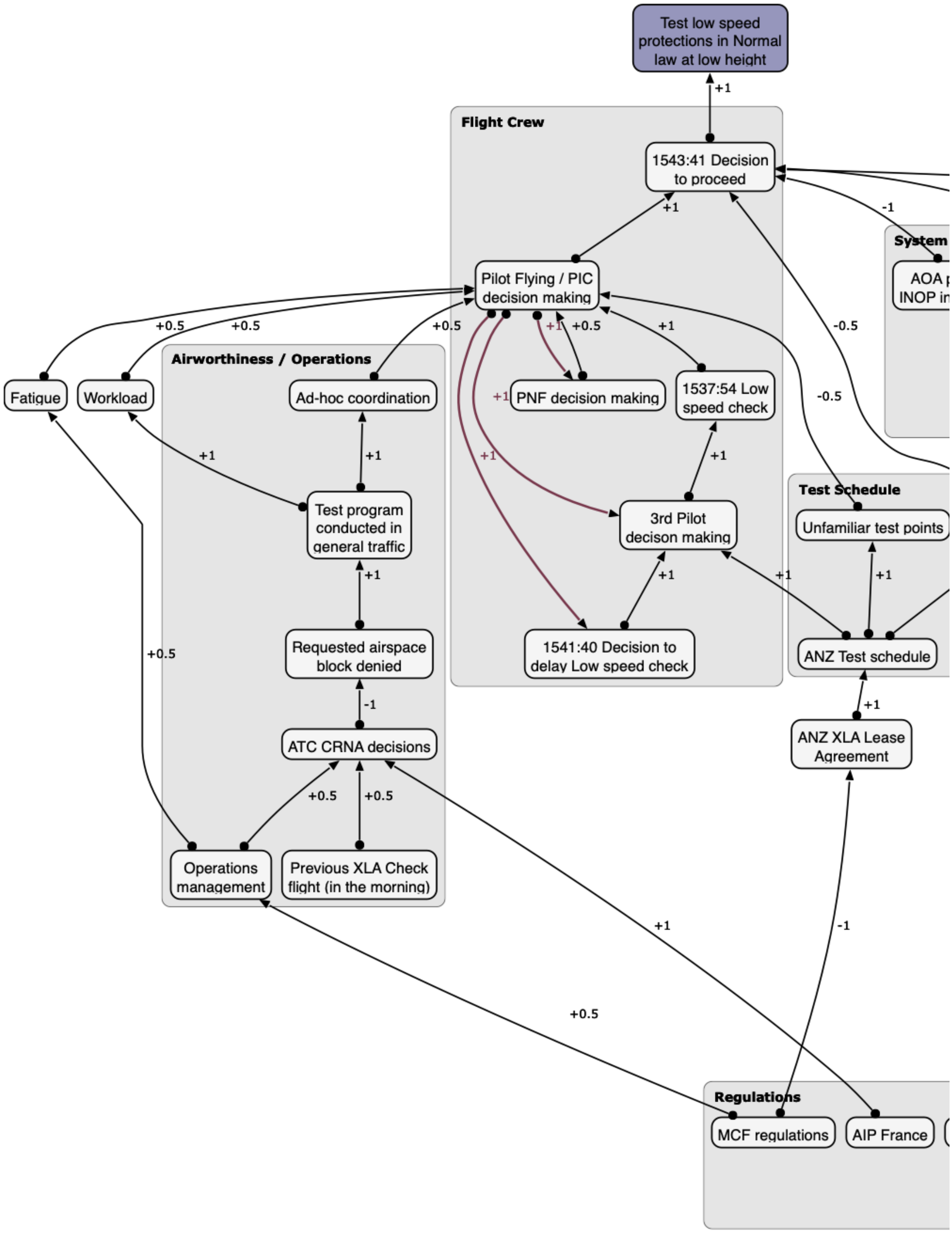
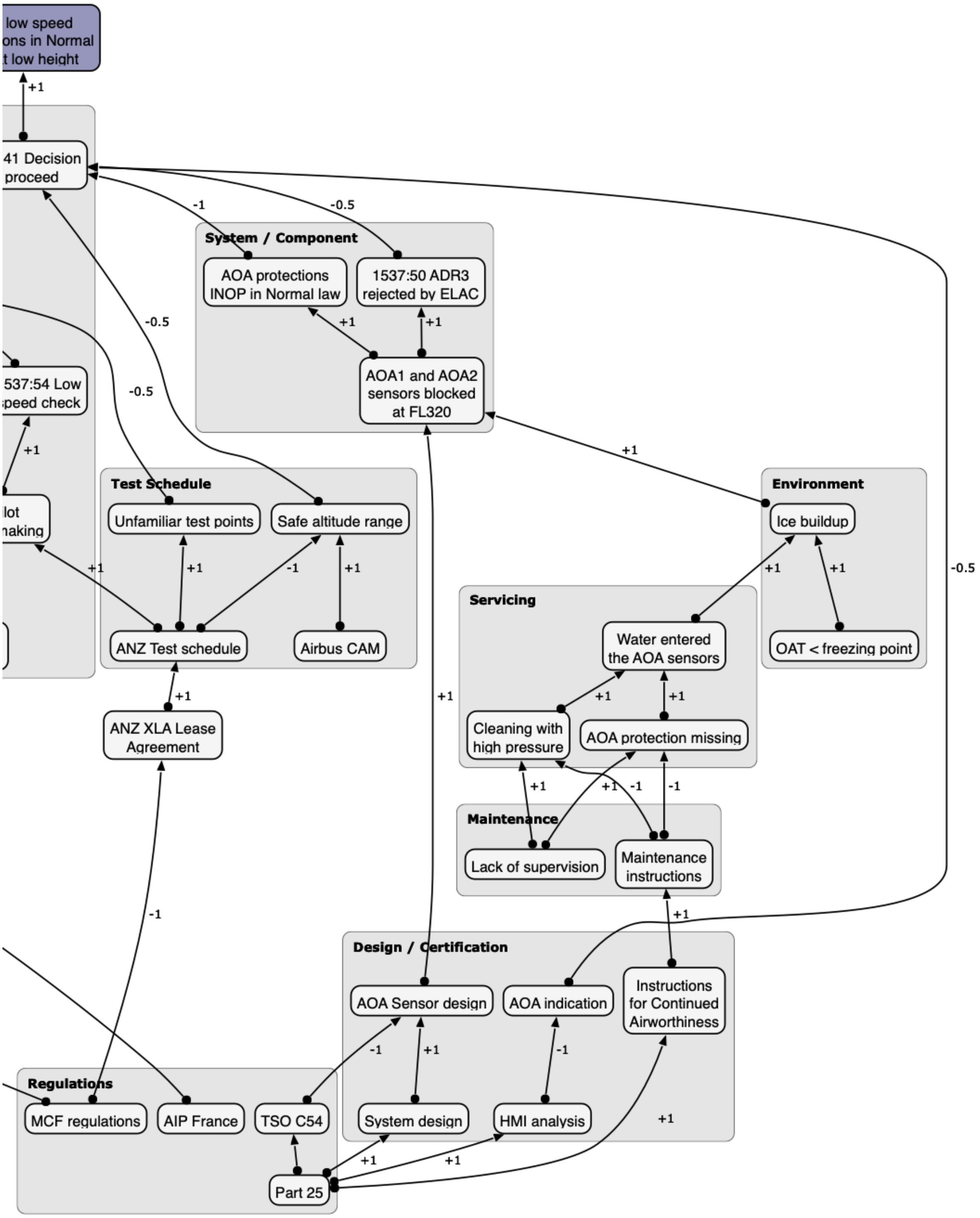


Figure 35

Pre-transition Concept map for 2008 Perpignan (cont.)



The ANZ pilot was probably the most experienced pilot for the purposes of the intended flight program. He was a training and check captain on multiple Airbus and Boeing types, and he was authorized by ANZ to conduct operational check flights. Prior to the accident flight, he was not involved in check flying an A320, though. He was formally designated as a cockpit observer for the accident flight and as the commander for the return flights from Frankfurt to New Zealand. Based on the CVR transcript, the ANZ pilot's role could be more accurately described as an additional third crew member, actively involved in decision making and directing the check flight program. He did not understand the German language, which was not a critical factor until the GXL crew repeatedly reverted to their native language when under time pressure (BEA 2010, pp. 23-24). In summary, during most of the accident flight, the PF relied on the ANZ pilot's experience and decisions when controlling the test sequence and safety parameters associated with the test points.

Ad-hoc Coordination. The French airspace sector controller's decision to reject non-standard manoeuvres requested by the accident crew while flying in shared airspace, and the PF's subsequent decision to continue with the programmed flight plan without a designated airspace block, led to the unintended consequence that all check flight test point had to be planned in an ad-hoc manner.

That decision, in turn, reinforced the ANZ pilot taking up a test engineer’s active role in directing the program, a role he was not trained for. This ad-hoc arrangement contradicted what was agreed to during the pre-flight briefing between the crew members and was likely a causal factor when the PF’s initial decision to postpone the low speed test from the program was not acknowledged by the ANZ pilot. To the contrary, the ANZ pilot repeated his request a short time later. The two captains’ exchange indicates that they did not share the same mental model about the risk involved in performing the low speed test during a critical flight phase (BEA, 2010, p. 161).

Test Schedule. To complicate the decision context even further, the ANZ pilot used an ANZ test schedule which was similar to the Airbus Customer Acceptance Manual (CAM), but lacking in detail, such as safe altitude ranges associated with in-flight checks. Furthermore, several test points were unfamiliar to the PF. CVR records suggest that the exact purpose and minimum safe entry and exit criteria for some elements in the test program may not have been fully understood by the crew members (BEA, 2010, pp. 125-163).

Design and Certification. The crew was not aware that minutes before the fateful decision to perform the low speed test, the automated software logic in the Elevator and Aileron Computers (ELAC) rejected the only valid air data reference unit. The valid AOA sensor was simply voted out by the certified system logic. Unbeknown to the crew, their AOA protections were inoperative in normal law. While the Bureau referred to the AOA sensor's inadequate design and certification process, the relevant systemic factors were not annotated as causal in the final report (BEA, 2010, pp. 183-186).

Decision to Perform the Low Speed Test at Low Height. The Bureau concluded that the crew's decision was a causal factor contributing to the accident. In hindsight, knowing that AOA sensors 1 and 2 were blocked and the remaining valid AOA value from 3 was rejected by the automated logic, preventing AOA protections to trigger during the low speed test, the Bureau's conclusion is defendable.

From the accident crew's perspective however, the decision to proceed with the test at a low height was made in a different context. They had no information about the blocked AOA sensors, and they did not know the accident outcome. Applying a counterfactual test, had they decided not to perform the test at a low height and delay the low speed test to a later segment where the Airbus recommended minimum safe altitude range of FL110 to FL140 was met, a safe outcome would not have been assured.

The minimum safe altitude range was prescribed for a crew complement with a flight test pilot monitoring and supervising the customer acceptance flight. It is highly unlikely that pilots not qualified for experimental test flights would have been able to recognise, and then recover from, the scenario faced by the accident crew where the flight envelope protections subject to the test point had failed in a silent manner.

6.2.5 Significant Events During the Transition to Recovery

At 1543:51, the PF asked the ANZ pilot for limit speed values applicable for the test point, who gave an ambiguous answer by "just… to come right back to alpha floor activation" (BEA 2010, p. 18).

At 1544:30, the PF stabilized the aircraft at 3,000 ft in a landing configuration. Flight directors were still active, and the PF changed the vertical mode V/S +0000. The speed was 136 kt.

At 1544:44, the A320 was at 2,980 ft and at a speed of 123.5 kts which corresponded to VLS. Thirteen seconds later, while near the LANET waypoint, a triple click can be heard on the CVR when the AP/FD lateral mode changed from NAV to HDG. One second later the aircraft was at 2,940 ft and a speed of 107 kts, corresponding to V_{min}.

By 1545:05, the horizontal stabilizer reached the electric pitch-up stop and did not move from that position until the end of the recording. The *stall warning* sounded when the recorded pitch angle reached 18.6 degrees. The airplane was at 2,910 ft and was flying at a speed of 99 kt (BEA, 2010, p. 18).

6.2.6 Significant Events During the Recovery Attempt

Seconds after the stall warning, the thrust control levers were moved to the *TO/GA position*, followed by a symmetrical increase in both engine N1 RPM values. The airplane started to roll to the left, however the PF countered the roll. The bank angle reached 8 degrees to the left and the speed was 92.5 kt. At 1545:11, the wings were straight then the aircraft began to roll to the right. The PF joystick was moved to the left stop, coordinated with left rudder pedal movement. At 1545:12, both flight directors disengaged, and the auto-thrust disarmed 2 seconds later (BEA, 2010, p. 18).

At 1545:15, the flight control laws moved from *normal to direct law*. The bank angle reached 50 degress to the right. The PF lateral input was still at the left stop. At the same time, the PF joystick reached the forward pitch-down stop position. The recorded pitch value was 11 degrees, speed at 100 kts and altitude remaining about 2,580 ft.

A 1545:19, the stall warning stopped. The elevators reached their maximum nose-down position. The bank angle reached 40 degrees to the left. One second later, the recorded pitch value was 7 degrees with wings close to level and speed at 138 kt, followed by pitch and altitude increasing. The airplane was climbing and airspeed dropping, which activated the stall warning again at 1545:36. Three seconds later, the PM retracted the landing gear (BEA, 2010, pp. 18-19).

At 1545:40, the flight control law passed to *abnormal attitudes*. The recorded pitch was 52 degrees nose up, the bank angle reached a maximum of 59 degrees to the left and the accelerometer recorded a value below 0.5 g. The PF roll input was practically neutral, while the pitch input was still forward but not constantly at the forward stop.

At 1545:42, the airspeed dropped below 40 kts and two seconds later the pitch was 57 degrees nose up at an altitude of 3,788 ft. The stall warning stopped, then started again about five seconds later. The airspeed started to increase, and the pitch value reached 7 degrees nose down. The airplane was banking left at 10 degrees. While the PF maintained lateral input to the left stop; the aircraft began to roll to the right, the bank angle increasing from 3 degrees to 97 degrees right. The pitch value reached 42 degrees nose down.

At 1545:58, the PM selected flaps and slats to position 1, then returned to position 0 only two seconds later. The PF made several inputs on the flight controls and thrust levers.

At 1546:00, the stall warning stopped, followed by a continuous repetitive chime of a Master Warning for two seconds. One second later, the airplane was 51 degrees nose-down at a bank angle of 45 degrees right, accelerating through 183 kts and remaining altitude at 1,620 ft. The PF pitch input remained at the rear stop from that moment (BEA, 2010, p. 19).

At 1546:04, an EGPWS "Terrain Terrain" warning was recorded by the CVR, followed one second later by another Master Warning. The flight recordings stopped at timestamp 1546:06.8Z. The last recorded pitch value was 14 degrees nose down, a right bank angle of 15 degrees, airspeed at 263 kts and an altitude of 340 ft (BEA, 2010, p. 20). Between the first stall warning and the timestamp of recordings stopped, there were sixty-two seconds.

6.2.7 What Systemic Factors Contributed to the Failed Recovery Attempt and Subsequent Safety Loss

The focus question in the second part of the case study is more aligned with the causal analysis performed by the Bureau. The main difference, as illustrated in Figure 36, is the addition of negative causal links, which guide the analyst's attention to causal factors which did not support the crew's decision making process or indicate interactions which may have triggered unexpected cascading failures propagating in the system.

The Bureau nominated the crew's incorrect management of the strong pitch up moment generated when the thrust vectors were suddenly increased by the crew during the initial stall recovery attempt as causal and contributing to the accident outcome. The Bureau also found that the crew failed to identify that the horizontal stabilizer reached the pitch-up stop position, did not correct the trim, and did not reduce engine thrust during the recovery (BEA, 2010, p. 98).

There is no intent to revisit the causal factors annotated by the Bureau. This analysis will focus on the differences between the BEA causal analysis and the causal patterns uncovered when accounting for negative causal relations.

The only remaining serviceable AOA sensor 3 activated the initial stall warning at 1545:05, which sounded 13 times. The stall warning was activated the second time at 1545:36 and sounded 12 times. The third and final stall warning series was activated at 1545:52 repeated 9 times (BEA, 2010, p. 164).

The airplane was flying at a low height above the sea and the PF applied a maximum thrust recovery technique, trying to accelerate his way out from the stalled condition by rapidly gaining airspeed. The A320 family is equipped with twin high bypass ratio jet engines, the nacelles and pylons integrated with the airframe in a below-the-wing configuration. In that conventional arrangement, a sudden increase in thrust, like the one applied by the PF during the accident flight, generates a strong pitch up moment on the airframe as the forward thrust vector acts below the airplane’s centre of gravity.

Figure 36

Post-transition Concept map for 2008 Perpignan (continued overleaf)

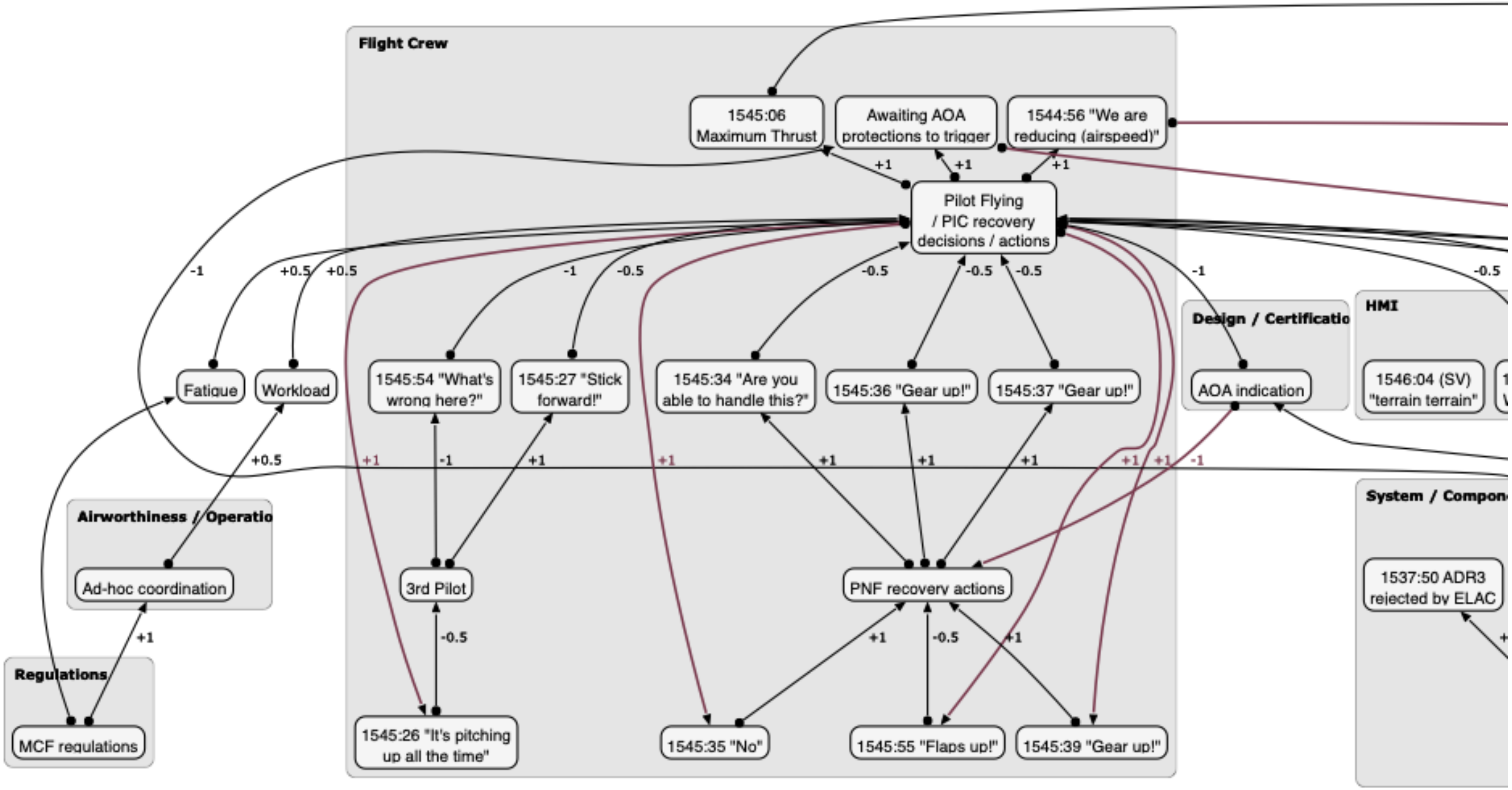
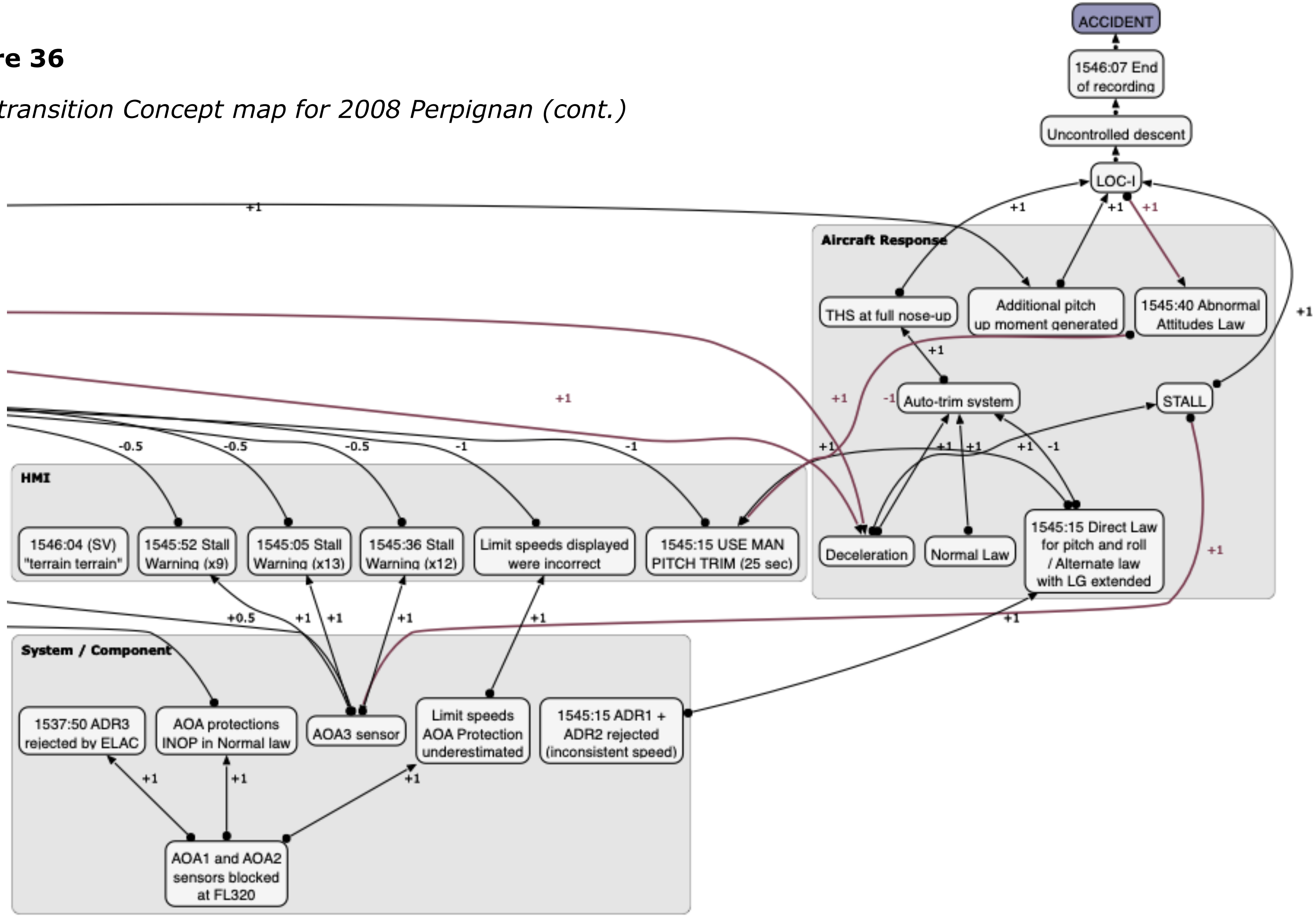


Figure 36

Post-transition Concept map for 2008 Perpignan (cont.)



The Bureau’s analysis that at 1545:15 a “use manual pitch trim” advisory appeared on the instrument cluster and remained visible for 25 seconds is correct, however, the final report failed to mention that the same advisory masked the fact that the auto-trim system was deactivated by an automated flight control law change. During that same period the first and second series of stall warnings were activated by the remaining warning loop (BEA 2010, p. 93).

Unbeknown to the crew, at 1545:15, ADR1 and ADR2 were rejected by the automated architecture due to inconsistent speed logic. At that moment, the flight control system degraded to direct law for pitch and roll (and alternate law when the landing gear was extended). This in turn meant that the auto-trim system was no longer functional and left the trimmable horizontal stabilizer at a full nose-up position. Recording the pilots’ inaction on the manual trim system in isolation, without mentioning the propagating effect of the deactivated auto-trim in a worst case (full nose-up) configuration, does not provide a full causal description (BEA, 2010, pp. 96-97).

The underlying systemic interaction missed by the Bureau’s final report is a critical finding for a highly automated fourth generation FBW design. For almost 8 minutes, starting when the correct ADR3 values were rejected by the software logic, the automation relied on incorrect air data fed by ADR1 and ADR2. During that whole period the AOA protection underestimated the limit speeds and those incorrect limit speeds were displayed on the primary flight displays. Later, the software logic caught up with ADR1 and ADR2, which led to all three ADR units being rejected and the incorrectly functioning flight envelope protections deactivated (BEA, 2010, pp. 183-186).

At 1545:26 the PF exclaimed that “Ja it’s pitching up all the time”, which correctly identified the aircraft response. The ANZ pilot responded by an instruction directed the PF when exclaiming “stick forward” (BEA 2010, p. 164). FDR records confirmed that the PF had repeatedly applied full or partial stick forward prior to that exchange. Both captains had correctly identified and focused on resolving the critical full nose-up pitch tendency of the airplane. Unfortunately, they did not have sufficient time to diagnose that the automated flight control system had commanded the horizontal stabilizer into a full nose-up trim position prior to the system logic suddenly removing the full authority of the auto-trim system.

6.3 Within Case Analysis

6.3.1 Introduction

The categories, concepts, and causal relationships identified in the pentavalent causal maps served as an input to the final part of the case study. As explained in the methodology chapter, the study aims to review basic patterns in how adaptive systems fail, namely decompensation, locally adaptive / globally maladaptive responses, and controllers working to flawed models or stuck in outdated behaviours (Woods & Branlat, 2017). To achieve that aim without a complex network model of adaptive units, the case study assumes a simple interaction model between the main controllers, including a) the crew, b) the automated systems, c) the MRO, d) GXL as a management unit, e) ANZ as a management unit, f) the ATC unit, g) Airbus design and airline support units, and h) the regulatory authorities (LBA, EASA, NZ CAA). The list is in order from lower to higher level adaptive units. The crew is the only adaptive unit where a refinement was necessary to differentiate between the PF, the PM, and the ANZ Observer (acting as test engineer).

Table 5 provides a high-level summary of the key events and basic (mal)adaptive patterns revealed from the pentavalent causal maps. The table introduces time as a critical dimension for failure analysis. Following the modified resilience terms introduced by Woods (2018), the analysis follows key challenges and surprises (events), the actions deployed (response), the risk of saturation (compensation), any potential misalignment or mismatch between the units (coordination), and any evidence of being stuck in outdated models when the environment has changed, or a new disturbance was introduced (revised model).

Table 5

Basic Patterns in Adaptive System Failures – Case B

Time	Key Event	Response	Compensation	Coordination	Revised model
0930Z	GXL B737 check flight	ATC sectors	Flight path changes		Updated model
1430Z	Maintenance release	Crew decision GO	Control transfer	Mismatch	Outdated model
1433Z	Delayed departure	Crew decision GO	Crew not at optimum	Potential mismatch	Outdated plan
1452:28Z	ATC rejected test block	Crew decision GO		Ad-hoc	Updated model
1454:25Z	Delay FLT CTRL checks (Normal law)	Crew decision GO	Ad-hoc testing		Outdated model
1503Z	AOA values frozen	Automation	Silent transfer	Mismatch	
1533:22Z	FLT CTRL test (Alternate) request	Crew decision GO	Below minimum safe altitude	Mismatch	Outdated plan
1536:47Z	FLT CTRL test (Alternate)	Crew decision GO	Test c/o at FL120		Outdated plan
1537:52Z	Low speed test request	Observer	Pilot Flying / Ad-hoc planning	Cross-purposes	Outdated plan
1541Z	PF suggested delaying low speed test	Pilot Flying	Observer did not confirm plan	Cross-purposes	Outdated plan
1543:37Z	Autopilot disengaged	Observer	Ad-hoc testing	Cross-purposes	Outdated plan
1543:48Z	Low speed test	Crew decision GO	PF is the fallback	Mismatch	Outdated plan
1545:05Z	Stab reached electric pitch-up stop	Automation	Airframe	Mismatch	Outdated model
1545:05Z	Stall warnings start	Pilot / Thrust	Thrust set TO/GA		Updated model
1545:12Z	Both FD and auto-thrust disengaged	Automation	Control transfer	Potential mismatch	Updated model
1545:15Z	FLT CTRL laws from	Automation	PF inputs	Mismatch	Outdated model

Time	Key Event	Response	Compensation	Coordination	Revised model
	normal to direct				
1545:15Z	Use Manual Pitch Trim	Automation	Silent transfer	Mismatch	Outdated model
1545:30Z	"Alpha floor"	ANZ pilot	Saturation	Gap	Stuck
1545:34Z	Are you able to handle this?	GXL Pilots	Decompensation	Potential mismatch	Stuck
1545:40Z	FLT CTRL law in abnormal attitudes	Automation	Saturated	Mismatch	Stuck
1546:04Z	EGPWS Terrain Terrain	Automation	Decompensation		Updated model

6.3.2 Adaptive System Failures

Communication and coordination – the first adaptive failures occurred during the strip and paint operations of the airframe. Chemical stripping was not fully effective, and the fuselage had to be sanded during the layover. A series of communication failures between the MRO supervisors, the painters, GXL representatives, ambiguous Airbus instructions and EASA rules, resulted in the MRO releasing an airplane with water present in AOA sensors 1 and 2. Unknown to the crew, a conditional surprise was transferred from the MRO to the automated systems and the pilots (BEA, 2010, pp. 69-70).

Misalignment – the delay put considerable pressure on an already fatigued GXL crew. During the flight, it became evident that there was a misalignment between the ANZ pilot's and the GXL crew's objectives regarding what test points can be achieved prior to returning to Perpignan airport and what needs to be postponed to the remaining Frankfurt leg. The ANZ pilot shifted roles from a designated observer to a test engineer capacity (compensation) in an unusual crew complement.

Local adaptations – an earlier GXL B737 check flight that same morning required extensive coordination between relevant ATC sectors. When the accident crew requested the same adaptive response (compensation), the ATC unit rejected the request (potentially at risk of saturation). The crew's updated mental model about the airspace environment offered two alternatives: a) abandoning the check flight program, or b) proceeding on an opportunity basis (ad-hoc coordination).

Outdated plans – flight control tests in normal and alternate law had to be delayed meeting ad-hoc coordination demands. Several test points were performed below recommended minimum safe altitudes. By this point, the crew was significantly behind the original test schedule (BEA, 2010, pp. 16-17).

Silent reconfiguration – unbeknown to the crew, the highly-automated airliner rejected the only remaining valid AOA 3 source and reconfigured the flight envelope protections in line with the designer's envisaged failure scenario, software logic, and remaining components (not condemned) in the system architecture. From that point, there was a complete mismatch between the crew’s mental model and the automation’s working model about remaining air data sources and envelope protections (BEA, 2010, pp. 183-186).

Working at cross-purposes – when the ANZ observer requested the low speed test during the Perpignan approach phase, the GXL and ANZ pilots started working at cross-purposes to each other. The GXL captain’s primary objective was to conduct a safe approach to Perpignan. The ANZ pilot was preoccupied with completing the remaining part of the test schedule.

Decompensation – the remainder of the accident flight can be best characterized in terms of the increasing saturation experienced by the main controllers, i.e. the automation and the crew. There was a growing gap between the pilots’ mental model (outdated and incorrect) and the automation’s working image of the situation (incorrect). The GXL pilots were likely approaching full saturation when they reverted to their native German language. Upon being queried by the PM, the PF admitted that he cannot handle the airplane upset. A few seconds later the automation reached the limit of its control authority when flight control laws reverted to abnormal attitudes (BEA, 2010, p. 164).

Outdated models and behaviours – the analysis also revealed a common adaptive trap for the three main controllers. For example, the ANZ pilot’s mental model was flawed when he exclaimed “alpha floor, we are in manual”, not realizing that the flight control laws had passed to direct law 15 seconds earlier, removing any flight envelope protections (BEA, 2010, p. 164). He did not understand German which created a critical communication gap between the crew members. Shortly after the ANZ pilot, the GXL pilots, and the automated systems appear to have been overcome by the magnitude and pace of disturbances, all stuck in outdated control models.

Flight deck effects – during the investigation Airbus highlighted that the pilots failed to detect several anomalies indicated in the cockpit. When those subtle coordination attempts are situated in a real-life (or flight simulator) setting, it is very difficult to substantiate the designer’s view. The rejection of AOA 3 was only indirectly indicated by the loss of CAT 3 DUAL approach capability, the erroneous limit speed calculations were only reflected in a CHECK GW message on the MCDU, and an amber USE MAN PITCH TRIM flag was displayed on the two PFDs for 25 seconds. In real life, the crew had no capacity to analyse the situation, having been overwhelmed by the rapid pace of events and warnings (BEA, 2010, p. 92-93).

Time – the airplane impacted the sea sixty-two seconds after the initial stall warning at 1545:05Z. The PF recognised the initial stall and attempted recovery by applying TO/GA thrust and pitching the nose down, not realizing that he could not overcome the pitch up moment generated by a stabilizer that was moved to the pitch up stop during deceleration (BEA, 2010, p. 86). A series of adaptive failures prevented the crew from correctly analysing the loss of control scenario under extreme time pressure. It follows that the lack of context-sensitive and salient warnings about the incorrect stabilizer trim position and actual AOA values were likely causal in the failed recovery attempt.

6.4 Discussion of the Findings

Focus Question One: Why did it make sense for the crew to continue the test program?

In summary, all accident crew members were very experienced A320 operators, the GXL pilots in routine line operations, supported by an ANZ pilot with some check flying skills. The unusual crew composition led to the ANZ pilot acting in an informal test engineer capacity. A series of adaptive traps led to them working at cross-purposes with each other, culminating in a tacit approval to perform an ad-hoc test of flight envelope protections below the recommended minimum safe altitude. Unknown to them, the automated system logic masked critical component failures and removed low speed flight envelope protections earlier in the test sequence.

Focus Question Two: Why did the recovery attempt work / fail to work?

The case study revealed basic adaptive system failures and patterns when the primary control units interacted with each other, including the a) the pilots, b) the automated systems, c) the GXL and ANZ management units, d) the ATC unit, e) the MRO, f) the Airbus design office, and g) the regulatory oversight units. The unforeseen failure scenario in a triple redundant design architecture, coupled with local adaptations during design and operations, resulted in a brittle and globally maladaptive system response.

While the GXL captain, acting as pilot flying, was a very experienced management pilot, he did not have sufficient exposure to non-routine flying. By contrast, the ANZ pilot had considerably more check flying experience, neither his observer, nor his assumed test engineer role permitted him to support the recovery attempt in a fully effective capacity. The crew faced an airplane upset scenario where critical flight envelope protections suddenly disappeared in an almost silent fashion, in a way that was not foreseen during the design phase. The nature of their mission suddenly changed from a check flight to an experimental flight. None of the crew members were trained for such a scenario, nor were they required to be at the time.

Neither the automated flight systems, nor the flight deck interface provided a supportive decision making environment. The automation “knew” about the cascading failures and responded to those failures according to a pre-determined outcome between the remaining hardware architecture and software voting logic. The software logic was not a failure, it simply worked as coded. The lack of context-sensitive and salient warnings about the incorrect stabilizer trim position and actual AOA values were likely causal in the failed recovery attempt.

While the crew's decision to perform the low-speed protections test at a low height was suboptimal, there is no evidence to suggest that a FL110 to FL140 minimum safe altitude range prescribed in the Airbus CAM would have provided a sufficient margin for airline pilots to correctly diagnose an unexpected and novel airplane upset scenario (BEA, 2010, p. 59). The final report did not explore the question of minimum test point entry and exit safety criteria for airline check flight operations, but the underlying assumption that CAM safety margins would directly apply here appears to be incorrect.

Chapter 7 – CASE C: B737 MANUAL REVERSION CHECKS

7.1 Introduction

Event: Serious Incident – Boeing 737-73V, G-EZJK

Location: West of Norwich, Norfolk – United Kingdom

Date: 12 January 2009

Agency: AAIB

On 12 Jan 2009, a Boeing 737-700 operated by easyJet experienced a serious in-flight upset and loss of control incident in the vicinity of Norwich, United Kingdom. The airplane operated a combined check flight and customer demonstration program, having just completed a maintenance visit. During a flight control manual reversion check, the airplane violently pitched down and lost approximately 9,000 ft before the pilot was able to recover and land the airplane safely. There were two company pilots and two observers on board. None of the occupants suffered any injuries during the incident flight.

The AAIB's (2010b, p. 30) final report concluded that a wide range of systemic causal factors contributed to the serious in-flight incident. Only the more traditional and simple root cause type description was extracted from the report:

> At a simplistic level the sequence of events leading to the in-flight incident can be directly attributed to the wording of the customer request form, which recorded the aircraft was out of trim in the nose-down direction rather than the nose-up direction identified by the pilot.
>
> In this incident the circumstances which initiated the sequence of events can be traced to the fact that the pre-maintenance delivery shakedown flight was not adequately planned, controlled, or communicated between the operator and the MROs.

The comprehensive AAIB (2010b) investigation uncovered additional systemic contributing factors, for example the test schedules used by the crew, pilot training standards for check flying, operations management, design considerations and aircraft response, maintenance issues, and concerns about airworthiness oversight. For the sake of brevity, only a summary is provided in the following paragraphs, in lieu of directly quoting from the report.

For customer demonstration flights the manufacturers are reluctant to provide tests schedules due to their perceived legal liability. In this instance, the operator was using an uncontrolled and out-of-date document, which led to cockpit switching that left the rudder unpowered and subsequent rolling corrections reliant on the ailerons alone. The anomalies with the customer acceptance schedule would not have been a factor, but the commander decided not to use the AMM instructions (AAIB, 2010b, p. 15).

Flight crew selection and training: the commander was not a qualified test pilot and lacked the depth of understanding to devise his own check flight program. The copilot had only participated in a single check flight prior to the incident and never received any training in addition to his line flying duties. He was briefed before the incident flight during a taxi ride to the airport (AAIB, 2010b, pp. 22-23).

Operations management and oversight: the AAIB concluded that the operator's perception of the commander's B737 experience and technical knowledge led to a situation whereby the commander was relied on to oversee non-routine operations across the operator's large B737 fleet on his own. The lack of oversight was extended to the UK CAA, which did not consider check flight operations as part of their remit (AAIB, 2010, p. 26).

Conduct of the flight: various elements of the incident flight reflected practices that would have been unacceptable during line operations. The commander (PF) decided to handle the controls, navigation, and radio communications, allowing his copilot to focus on managing the test schedule. A maintenance engineer was observing the flight in the cockpit without a secure observer seat for the whole duration of the flight (AAIB, 2010b, p. 26).

Aircraft response: The commander could not apply sufficient force to the elevator controls to overcome the aerodynamic loads generated because of the balance tab position and the high speed of the aircraft. Had no other factors assisted in the recovery, the commander’s actions alone would not have been sufficient to prevent the continued descent (AAIB, 2010b, p. 27).

Systemic maintenance issues: the AAIB investigation uncovered a fragmented and incoherent maintenance organisational structure in charge of the redelivery program. Key stakeholders acted in silos; responsibilities defined by self-assessed boundaries. The lack of coordination and poor communication resulted in a single human factors issue propagating through the whole maintenance organisation, and a near loss of the airplane and the incident crew members. The UK CAA oversight was not sufficiently informed about the process to have identified the systemic issues (AAIB, 2010b, pp. 28-29).

Specific maintenance issues: the incorrect transcription of the pre-delivery pilot handover was not an uncommon mistake. The AAIB (2010b, p. 30) identified several similar mis-rigging incidents in the past, which were the result of incorrect maintenance records. The design induced inability of maintenance engineers to physically identify a mis-rigged flight control tab was a common occurrence. The AAIB (2010b, pp. 30-31) found that the B737 AMM did not provide sufficient guidance in performing this flight safety critical task. Such critical flight control errors will result in more severe aircraft upset incidents, unless a more robust system is in place to support subsequent check flight testing.

Continuing airworthiness requirements: the AAIB (2010b, pp. 31-32) found that this is an industry-wide safety issue, reflected in several investigations around the globe. While EASA rules required airline operators to demonstrate and local CAAs to monitor the continuing airworthiness of their aircraft, the lack of regulatory guidance meant that previously effective methods were abandoned and created a vulnerability for non-routine airline operations.

In an earlier phase of this research project, a text analysis was conducted for each check flight investigation report within scope (see chapter three and four). Figure 37 provides a graphical summary of the causal logic embedded in the results chapter. A causal mapping study delivered a different graphical representation for the same investigation results. Figure 23 illustrates the main causal categories extracted from the AAIB (2010b) report, along with the causal links and the relative strength of those causal findings, indicated by labels assigned to the directed edges between causal nodes.

Figure 37

Text Analysis for 2009 Norwich

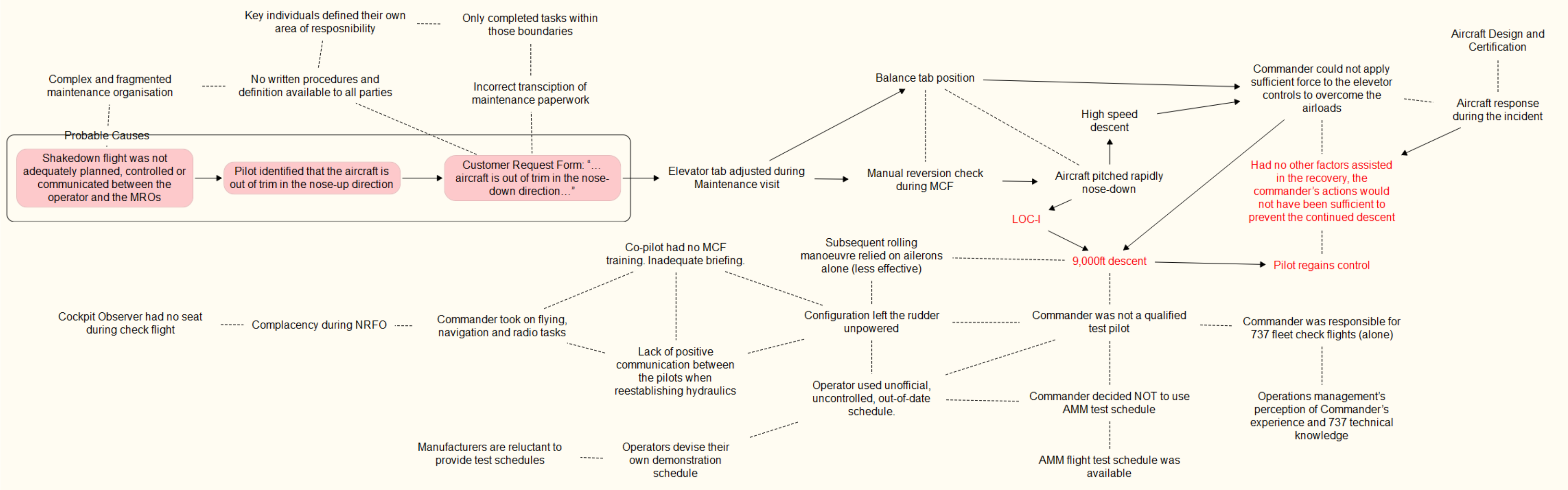
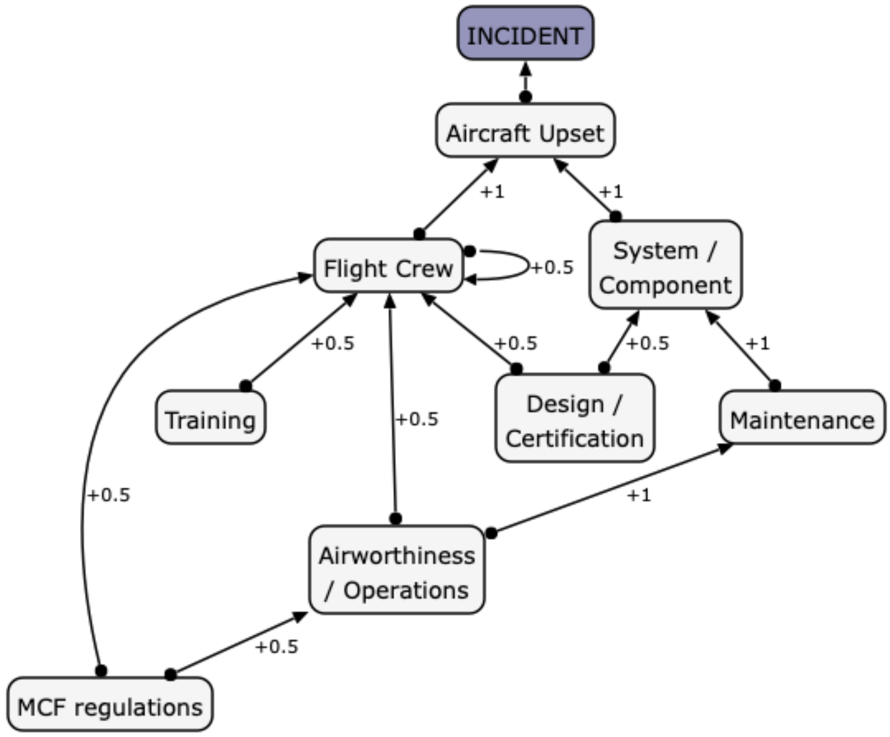


Figure 23

I12 – 2009 Norwich (duplicated for ease of reference)



As discussed in the previous chapter, it is inevitable that at a higher level of abstraction certain causal details remain hidden. When the two diagrams are compared, a simple linear translation can reveal those hidden details. For example, the causal statement that "the pre-maintenance delivery shakedown flight was not adequately planned, controlled or communicated between the operator and the MROs" translates to the causal link between the [Airworthiness / Operations] and [Maintenance] nodes, and the {+1} causal strength (immediate cause) assigned to the edge between those nodes (AAIB 2010b, p. 30).

The Civil Aviation (Investigation of Air Accidents and Incidents) Regulations (2018) give effect to Annex 13 of the Chicago Convention, establishing the AAIB, and the fundamental principles for investigating reported civil aviation accidents in the United Kingdom. The sole objective of an AAIB investigation is the prevention of accidents and incidents without apportioning blame or legal liability. AAIB investigation reports record a sequence of events, provide analysis of factual evidence, then list safety issues, adopt safety recommendations, and record safety action, as relevant.

The case study relies on the final report issued by the Branch, albeit with a different scope, when seeking answers to the modified focus questions. In this instance, the case study is guided by a) reasons why it made sense for the accident crew to proceed with the manual reversion check of the elevators, and b) reasons why the recovery attempt was successful. There is no intent to criticize the AAIB (2010b) investigation process or their findings. The sole focus of this case study is to revisit the incident flight in the context of this research project, especially the safety recommendations adopted by the Branch in their final report.

7.2 Data Analysis

7.2.1 Nominated Transition Point

Based on the sequence of events and evidence available from transcripts in the AAIB (2010b) final report, the commander's decision to give the go ahead for putting the aircraft in a manual reversion mode at 1536:47Z was nominated as the point where the check flight sequence started transitioning from the original test program objective to a recovery phase (p. 4).

The case study describes noteworthy events leading up to the accident flight and significant events pre- and post-transition. The pre-transition narrative focuses on factors contributing to the commander's decision to proceed with the manual reversion test. The post-transition narrative revisits available evidence in the context of the successful recovery attempt, including the key contributing factors that resulted in regaining control and a safe landing.

7.2.2 Noteworthy Events Prior to the Incident Flight

In December 2008, the incident B737 aircraft had been inducted to a Part 145 MRO at Southend Airport for scheduled maintenance and bridging checks, in preparation for handing the aircraft back to its owner at the end of the agreed lease period. Post-maintenance, a combined check and demonstration flight was required in accordance with the Customer Demonstration Flight Schedule (CDFS) agreed between the airline and the owners. It was the commander of the incident flight who ferried the airplane to Southend. During that ferry flight he had carried out a series of in-flight checks, annotating incoming defects on a printed copy of the CDFS (AAIB, 2010b, p. 15).

On 12 January 2009, the commander returned to Southend for conducting the post-maintenance check flight. Before the flight, the commander received a verbal brief from the MRO crew chief about the maintenance work package that had been carried out during the layover. In his post-incident interview, the commander recalled being told about the elevator balance tab adjustments and was given pages extracted from the AMM to assist him in conducting an elevator power-off test (manual reversion check) and in identifying any asymmetrical flight control forces. The B737 AMM prescribes that the in-flight testing must be completed before the airplane is returned to revenue service (AAIB, 2010b, p. 2).

Prior to the 1400Z departure, the commander reviewed technical log entries and confirmed the ATC arrangements for the check flight to be conducted in restricted military training airspace. The commander and his first officer copilot were joined by two technical observers who were representing the owner and the receiving airline operator. The pre-flight checks were completed on the technical apron with no fault found and the aircraft left the ramp with the commander acting as PF (AAIB, 2010b, p. 2).

7.2.3 Significant Events Leading up to the Transition Point

The commander conducted a series of checks at FL410 in a 45-minute window, then requested clearance to descend to FL150 where an APU bleed test was performed. Next, the aircraft was configured for the flight control manual reversion check. The aircraft was flying at FL150 at 250 kts. The fuel load was balanced, the autopilots were off, the stabilizer trim main electrical (STAB TRIM MAIN ELEC) and autopilot trim switches were set to cutout, and the aircraft was within trim limits. The CDFS required that spoiler A and B switches were selected off. It is noteworthy that all in-flight check items were configured and conducted in accordance with the operator’s CDFS and not the AMM extracts given to the commander by the MRO crew chief (AAIB, 2010b, p. 2).

Before the manual reversion check began, the individual hydraulic systems were isolated in turn. First, FLT CTRL B switch was set to the OFF position and the flight controls moved slightly. The switch was reinstated, then FLT CTRL A was put to the OFF position and the flight controls moved again slightly. The objective of alternating the switches was to check the flight controls for normal operation on a single hydraulic system. Operation was confirmed as satisfactory on both systems (AAIB, 2010b, p. 2).

At CVR timestamp 1536:44, the copilot confirmed that FLT CTRL A remained OFF and requested confirmation from the commander to proceed with de-powering the remaining hydraulic system. The commander responded with a short “Yes, go ahead” (AAIB, 2010b, p. 4). The commander then released the controls and the copilot selected FLT

CTRL B to the OFF position, removing all hydraulic assistance (A + B) from the primary flight controls at 1536:47.

7.2.4 Why Did It Make Sense to the Commander to Proceed with the Elevator Power Off Test

In hindsight, knowing the consequences of the commander's decision to proceed with the manual reversion check in a CDFS configuration, it would be easy to highlight that the EZY flight crew should not have attempted the elevator power off tests on 12 January 2009. As found by the AAIB (2010b, p. 25), that decision triggered an unexpected aircraft response, which then resulted in the commander losing control of the airplane and rapidly accelerating toward the terrain below. Framing the safety issue in that way, the analysis relies on knowing the accident outcome, which is counter-productive and carries the risk of introducing considerable hindsight bias during the investigation process.

A more interesting question is exploring the reasons why it made sense for the EZY commander to proceed with the manual reversion check. This section of the case study revisits that critical decision, evaluating the decision context from the crew's perspective. Figure 38 offers a graphical summary of systemic factors leading to the transition point.

Figure 38

Pre-transition for 2009 Norwich (continued overleaf)

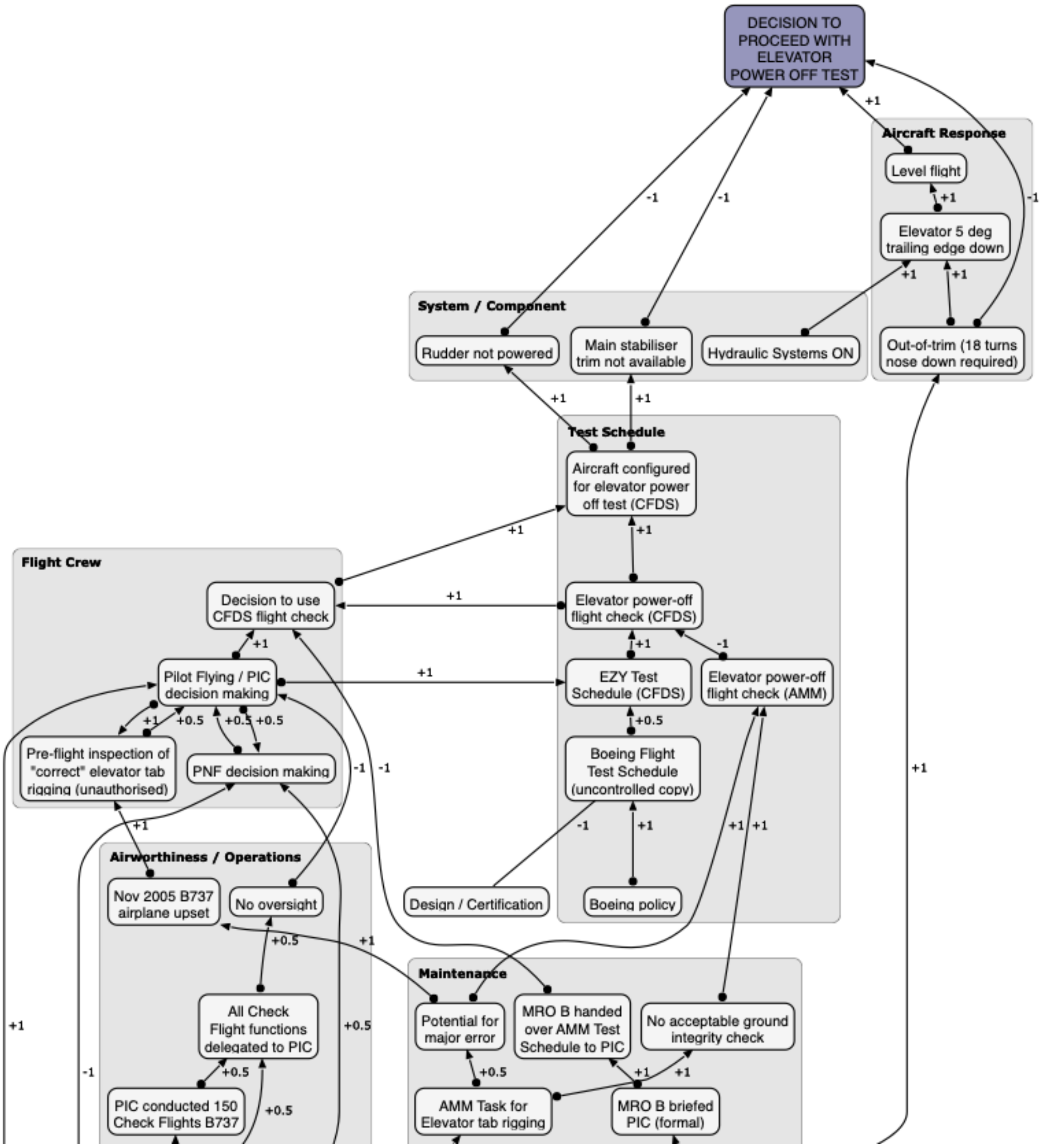
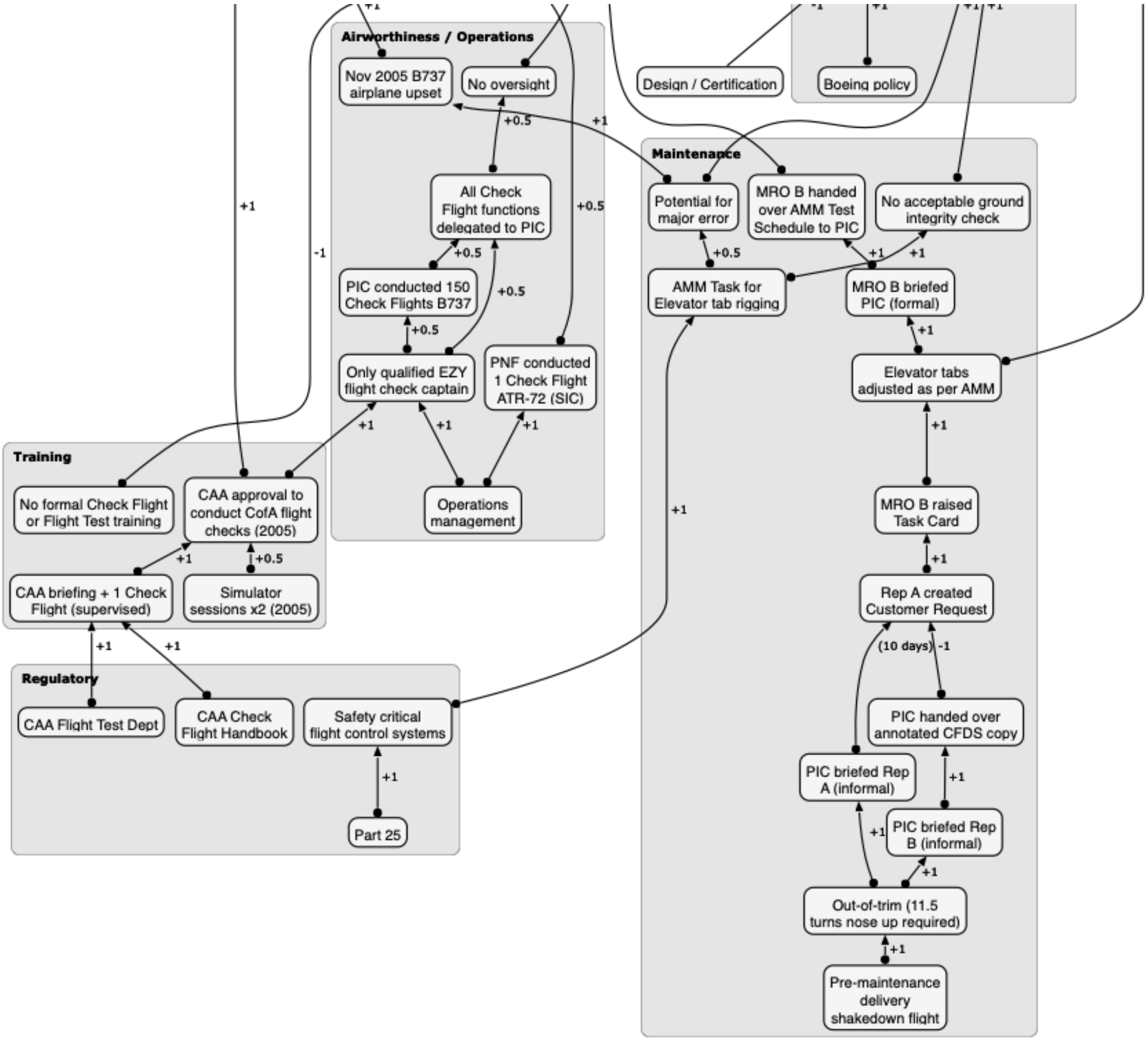


Figure 38

Pre-transition for 2009 Norwich (cont.)



Crew Experience and Composition. The PF was the only qualified and authorized EZY captain for conducting B737 check flights. Considering the size of the EZY B737 fleet at the time, it is not surprising that he had accumulated significant check flying experience, amassing a record of commanding over 150 check flights prior to the incident flight.

In 2005, the PIC received UK CAA approval to conduct airworthiness flight checks, having completed two dedicated simulator sessions, a CAA briefing, and one check flight under supervision of a CAA test pilot. Later that year, he received commendations from the CAA, when he successfully recovered an earlier B737 model EZY aircraft during a similar in-flight aircraft upset scenario, induced by an incorrect elevator tab adjustment during maintenance. In sharp contrast to the PIC, the first officer (PM) had only participated in a single ATR-72 check flight as second-in-command. He received no formal check or flight test training during his career (AAIB, 2010b, pp. 22-23).

The PIC was one of the most experienced B737 airline check pilots in Europe at the time of the serious incident and an expert in his field. At the same time, the AAIB (2010b, p. 17) raised concerns about the commander’s self-assessed technical competency, demonstrated by an unauthorized pre-flight inspection developed by the pilot to check the B737 elevator tabs for correct rigging during his pre-flight routine. Unfortunately, the B737 elevator tab design does not offer such an opportunity for a visual inspection to verify the correct flight control rigging when a person is standing on the apron.

CDFS Test Schedule. At the time of the serious incident, Boeing (and other airframe manufacturers) were reluctant to provide airline operators with a Flight Test Schedule (FTS), citing their potential legal liability. EZY obtained an unauthorized copy of the Boeing FTS during one of the factory acceptance flights and developed an in-house EZY test schedule, referred to as the CDFS in the AIIB (2010b) report (p. 23). It appears that the person transcribing the Boeing FTS to the EZY CDFS may not have been fully aware of the full intent and scope of the prescribed test points, such as the elevator power off tests. The CDFS required the check flight crew to configure the airplane in such a way that the rudder was not powered, and the main stabilizer trim was not available during the manual reversion check. The CDFS configuration contradicted the AMM instructions (AAIB, 2010b, p. 25).

The commander made the decision to use the CDFS configuration despite the MRO crew chief providing him with the relevant AMM pages extracted from a controlled publication. During his post-incident interview, the PF stated that he found the AMM pages to be confusing and written for maintenance technicians, not for check pilots (AAIB, 2010b, p. 25). His comments could be seen as rationalizing his decision in hindsight, but his statement is more widely echoed in the check flight community.

Although the Boeing AMM complies with FAA regulations for instructions for continued airworthiness, the AMM pages for a manual reversion check have always been considered as problematic, unless the requirements are properly coordinated between maintenance engineers and airline technical pilots.

Maintenance Error and Aircraft Response. Figure 39 also illustrates the complex web of systemic maintenance errors which resulted in the airplane being released from maintenance with the elevator tabs grossly out-of-trim. The AAIB (2010b, p. 18) investigation found that the incident airplane required 18 turns nose-down on the elevator trim wheel, which was a major maintenance-induced defect.

The incident crew had no way of knowing about the maintenance error prior to turning off the hydraulic systems in-flight. In this instance, the commander happened to be the pilot who delivered the airplane to the maintenance provider. His mental model relied on the fact that he briefed relevant maintenance representatives, albeit informally, about the need to adjust the elevator tabs in the correct sense. Furthermore, the maintenance error was masked from the operating crew, even when only a single hydraulic system was powered. The B737 design ensures that the elevator has sufficient aerodynamic authority to counteract an incorrectly rigged elevator tab, provided the airspeed range remains within limits considered to be normal during line operations.

7.2.5 Significant Events During the Recovery Attempt

At 1536:47 when the PM selected both FLT CONTROL switches (A and B) to the OFF position, all hydraulic boost was removed from the primary flight controls. The elevator rapidly moved to an 8-degree trailing edge down position and the airplane pitched nose-down. The PF applied considerable force by pulling back the control column, temporarily returning the elevators to a trimmed position, but was unable to hold the required control force. The aircraft was descending at a rate of 3,100 ft/min (AAIB, 2010b, p. 2).

The commander stated that at that moment he decided to reinstate all hydraulics and knock off the manual reversion check. He also stated that he had been trained not to reinstate the hydraulics immediately, rather the aircraft should be rolled to unload the pressure on the elevators and the control column released before the hydraulics are back on. He understood that without that technique the B737 airframe can be overstressed, or the pilot can sustain a serious injury (AAIB, 2010b, p. 2).

The PF rolled the aircraft left to 70° before releasing the controls and at 1537:04 called for the PM to get ready for re-engaging both flight control switches. The PM responded with “Say again”, to which the PF replied, “And back” (AAIB, 2010b, p. 5). The PM stated that he did not understand the command. The descent rate was now 6,000 ft / min and increasing, and the airspeed was at 270 kts and accelerating. The aircraft continued to roll to 91 degrees, the maximum recorded bank angle.

The CVR transcript indicates that there was significant confusion between the two pilots. The PF believed that hydraulic power had been restored. The PF reduced engine thrust and selected the speedbrakes, but the speedbrakes did not extend, as the spoilers had been selected OFF as part of setting up the CDFS configuration. The PF then rolled the wings level and attempted to arrest the rate of descent which had peaked at 20,000 ft/min. The aircraft was in a 30-degree nose-down pitch after the aircraft had been commanded to roll left. The maximum recorded airspeed was 429 kts / Mach 0.719, the maximum vertical acceleration was 1.6g and the minimum recorded altitude was 5,655 AMSL. The Mach trim was activated by the aircraft above M 0.615, helping the PF with additional pitch up commands (AAIB, 2010b, p. 5).

At 1537:20, the PF made a PAN PAN call to ATC at FL114. The CVR recorded an aural overspeed warning at the time of the radio call. There was no recorded communication between the crew members since the PF said, “And back”. Shortly after levelling at 7,000 ft, about 76 seconds after the previous request, the PF instructed the PM to put the flight control switches back on. The control forces returned to normal at 1538:27, when both hydraulic system A and B were reinstated. Fifteen seconds later, the PF made another radio transmission to ATC and cancelled the PAN (AAIB, 2010b, p. 3).

During the post-incident interview with the AAIB (2010b, p. 3), the commander stated that he had considered repeating the manual reversion test during the incident flight. Fortunately, he elected not to repeat the test, as he was concerned about flying a potentially overstressed airframe. He kept the airspeed below 250 kts and configured it for landing very early during the approach phase. The PF assessed that the aircraft and flight controls were operating normally. At 1606Z, the airplane landed at Southend without further incident.

7.2.6 What Systemic Factors Played a Role in the Recovery Attempts

As illustrated in Figure 39, there were two recovery attempts made by the pilot. Based on the sequence of events reported by the AAIB (2010b), the first attempt has failed due to a communication breakdown between the two pilots. The PM did not understand the commander's instruction and the PIC did not react to the PM's request for clarification. As a result, the PM did not restore hydraulic power to flight control systems A and B when instructed by the PF to do so. As highlighted earlier, more than 76 seconds passed between the first and the final (full) recovery attempt without any recorded exchange between the two pilots.

By following the causal links in Figure 39, the following immediate causes played a role in the initial (failed) recovery attempt: the elevators were grossly out-of-trim (maintenance-induced), main stabilizer trim was not available (CDFS configuration), roll control was effected by ailerons only (rudder was unpowered in CDFS configuration), the speedbrakes did not deploy (spoilers were off in CDFS configuration), and flight control systems A and B hydraulics were not restored prior to the attempt. The pre-flight decision to reject the AMM, and use the CDFS configuration instead, was a common causal factor for several immediate causes listed here.

Fortunately, the second recovery attempt was successful. The post-transition map identifies the following positive systemic interactions which aided the PIC/PF in arresting the descent:

- In Nov 2005, the same pilot was involved in a B737-300 check flight and he successfully recovered that airplane in a pitch-up LOC-I scenario. B737 models have an almost identical elevator control system and similar control feel. The PIC was able to draw on that experience when formulating a response during the incident flight.
- The fact that the manual reversion check was started at FL150 with sufficient altitude margin. FDR data indicates that more than 50 percent of that safety margin was used up by the time the rapid descent was arrested by the pilot.

- Finally, the Mach trim activating and providing additional pitch-up input to the elevators.

The confluence of the above-mentioned three systemic factors played a critical role in the successful recovery attempt. The first and second factors are largely controllable by proper planning, preparations, and training. In contrast, the positive contribution of the Mach trim was beyond the control of the pilot. The causal trigger for the Mach trim activating was the airspeed increasing beyond Mach 0.615 during the overspeed event (AAIB, 2010b, p. 5).

The pilot’s control input aimed at arresting the descent and the speed required for activating the Mach trim were at cross-purposes with each other. In this case, it was a fortunate (unplanned) interaction between those two factors, but the same positive outcome is not guaranteed for future events. Even if the airspeed exceeds the limit Mach value during a similar high-speed descent, any number of systemic failures can prevent the Mach trim from activating or providing the required elevator control input.

Figure 39

Post-transition Concept map for 2009 Norwich (continued overleaf)

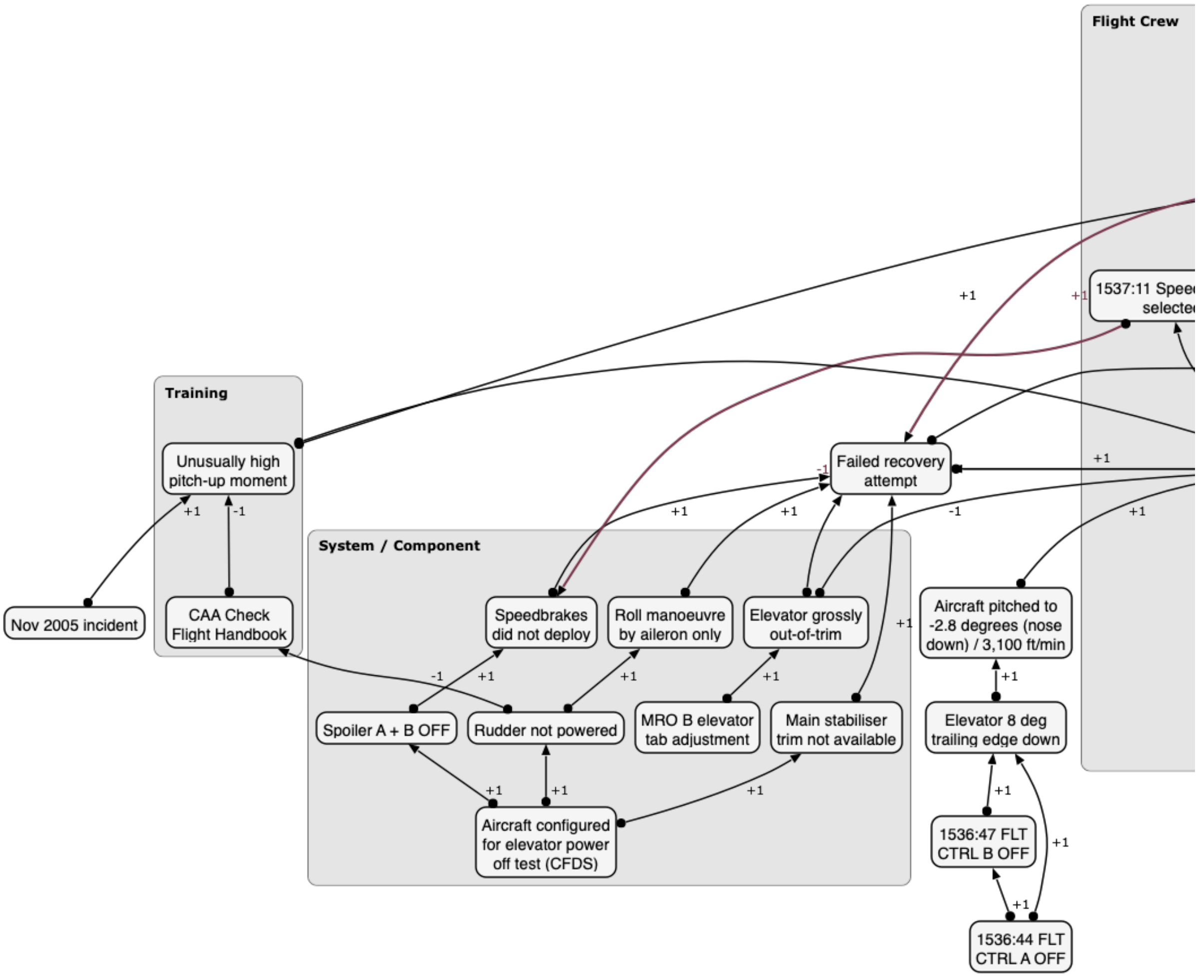
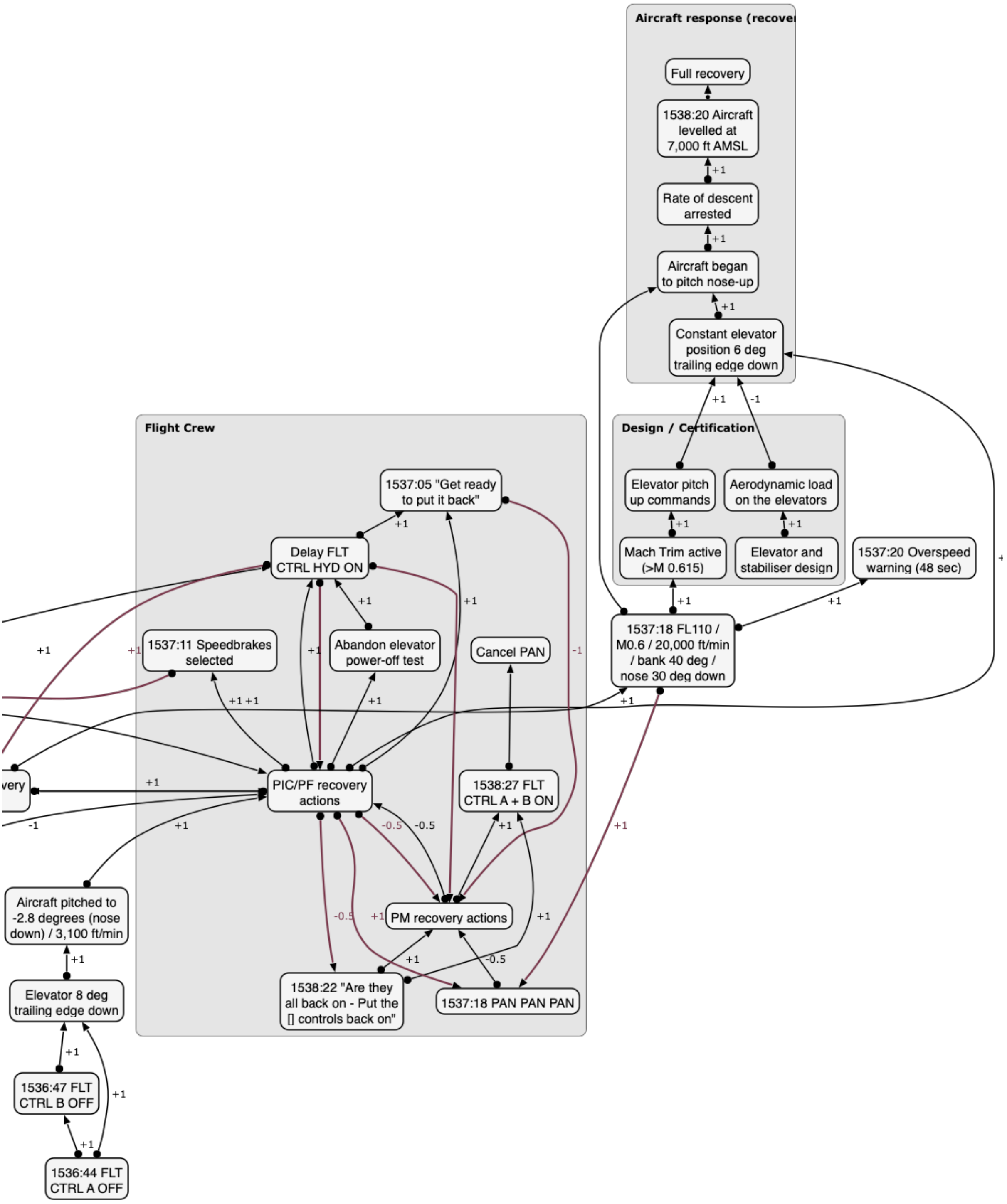


Figure 39

Post-transition Concept map for 2009 Norwich (cont.)



The AAIB (2010b, p. 27) observed that the PIC/PF could not apply sufficient control column force to overcome the aerodynamic forces generated during the high-speed descent. The final report also annotated that had no other factors assisted in the recovery, the pilot would not have been able to arrest the descent. The AAIB (2010b, p. 18) report dedicated a section to maintenance implications of the B737-700 design, but the report made no recommendations in a design context.

7.3 Within Case Analysis

7.3.1 Introduction

Building on the categories, concepts, and causal relationships uncovered in the pentavalent causal maps, this part of the case study elicits basic patterns observed in the data. As outlined in chapter three, the intent is to review the basic adaptive traps, including patterns of decompensation, locally adaptive / globally maladaptive responses, or controllers following outdated models or behaviours (Woods & Branlat, 2017).

To achieve that aim without a complex network model of adaptive units, the case study uses a simple interaction model between main controllers, including a) crew members, b) the automated flight control systems, c) the MRO units involved, d) the airline’s management unit, e) f) Boeing’s design and airline support organisation, and g) the regulatory authorities (UK CAA and EASA). The list is in order from lower to higher level adaptive units. For the crew and MRO units it was necessary to refine the model by differentiating between the PIC/PF, the PM, and MRO A and B, respectively.

Table 6 provides a high-level summary of the key events and basic (mal)adaptive patterns extracted from the pentavalent causal maps. The table introduces time as a critical dimension for failure analysis. Following the modified resilience terms introduced by Woods (2018), the analysis follows key challenges and surprises (events), the actions deployed (response), the risk of saturation (decompensation), any potential misalignment or mismatch between the units (coordination), and any evidence of being stuck in outdated models when the environment has changed, or a new disturbance was introduced (revised models).

Table 6

Basic Patterns in Adaptive System Failures – Case C

Time	Key Event	Response	Compensation	Coordination	Revised model
NOV2005	Uncommanded pitch-up during check flight	Pilot Flying	Control transfer		Updated model
01DEC2008	Maintenance delivery flight	Pilot Flying	Control transfer	Mismatch	Updated model
11DEC2008	Tech Log entry	MRO A	Control transfer	Mismatch	Outdated model
12JAN2009	Verbal briefing	MRO B	Control transfer	Mismatch	Outdated model
Pre-flight	Decision to reject AMM test schedule	Pilot Flying		Mismatch	Outdated plan

Time	Key Event	Response	Compensation	Coordination	Revised model
1400Z	Departure	Crew decision GO		Mismatch	Outdated plan
1535:40Z	FLT CTRL B OFF	Automation	Masking	Mismatch	Outdated model
1536:15Z	FLT CTRL A OFF	Automation	Masking	Mismatch	Outdated model
1536:47Z	FLT CTRL A+B OFF	Automation	Rapid ELEV movement		Updated model
1536:48Z	Recognition	Pilot Flying	Manual control		Updated model
1537:04Z	Verbal command to PM	Pilot Flying	Manual control	Mismatch	Outdated model
1537:11Z	Speedbrake handle	Automation	Not powered	Mismatch	Outdated model
1537:15Z	20,000 ft descent rate	Pilot Flying	Manual control	Cross-purposes	
1537:20Z	Overspeed condition	Automation	Mach trim	Ad-hoc	
1537:20Z	PAN PAN PAN	Pilot Flying	Saturated		Updated model
1538:25Z	FLT CTRL A+B ON	Pilot Flying	Pilot Monitoring		Updated plan
1538:42Z	Back under control	Pilot Flying	Cancel PAN		Updated model
	PF considered repeating manual reversion test	Pilot Flying	Airspeed kept below 250 kt		Revised plan
1606Z	Airplane approach and landing	Pilot Flying	Automation		Updated model

7.3.2 Adaptive System Failures

Recognition – in Nov 2005, during a B737-300 manual reversion check the PIC experienced and recovered a LOC-I incident when a strong pitch-up moment was generated by an incorrectly adjusted elevator trim system. The incident also served to update the pilot's mental model about B737 response and controllability during airplane upsets (AAIB, 2010b, p. 17).

Communication and coordination – after the ferry flight, control was transferred between multiple units in a complex organisational network. Communication breakdowns resulted in contradicting mental models generated by the maintenance engineers (MRO A and B) and the pilot (AAIB, 2010b, pp. 15-16). The B737 design provides no opportunity to perform an acceptable integrity check on the ground, i.e. there are no means to adjust models once coordination breaks down.

Outdated models and plans – the PIC's decision to reject the AMM was based on an outdated model, that in turn was generated from a flawed CDFS test schedule. The CDFS was drafted by the airline, compensating for Boeing's reluctance in releasing check flight instructions to airline operators (AAIB, 2010b, p. 23). At the time of departure, the crew's mental model was at odds with both MRO B and Boeing, planning to conduct the manual reversion check according to an outdated plan.

Masking – the automated system response has successfully masked the fact that the elevator control system was grossly out-of-trim. Isolating individual hydraulic systems (B, then A) was not suitable for uncovering a common cause failure mode (elevator tabs) (AAIB, 2010b, p. 18). As a result, the crew entered the manual reversion test point with a flawed and outdated mental model.

Decompensation – the violent pitch-down moment generated by the incorrectly adjusted elevator balance tabs saturated the automated flight control system (AAIB, 2010b, p.2). The system decompensated by transferring control to the PIC/PF. After that surprise event, there was a growing coordination mismatch between the two pilots.

At risk of saturation – during the failed recovery, the coordination gap was exacerbated by a communication breakdown between the pilots (AAIB, 2010b, p. 5). Both the PIC and the PM were following isolated and outdated mental models. There was a mismatch between the automated controller and the PIC that prevented deploying the speedbrakes. The PIC was at risk of saturation, reflected in a PAN PAN declaration upon receiving an overspeed warning in the cockpit.

Local and global adaptations – the automated Mach trim response and the PIC were independently working towards local goals during the rapid descent phase (AAIB, 2010b, p.5). According to post-incident statements and FDR records, it is only by coincidence that local adaptations implemented by the PIC and the automated controller achieved a globally adaptive response.

Updated mental model – full recovery was only achieved after the airplane was back under control in level flight. The lower cognitive demands enabled the PIC to fully scan his cockpit indications, only to realize that FLT CTRL A and B switches remained off during the recovery attempts. He formulated an updated plan and cancelled the emergency.

Residual misalignment – the PIC considered repeating the manual reversion test prior to returning to the maintenance base. During the post-incident interview, he stated that he elected not to proceed with that plan due to concerns about a potentially overstressed airframe (AAIB, 2010b, p. 3). His updated model was not fully aligned with the actual system state, an adaptive trap that was narrowly avoided in this instance.

Time – the crew was under considerable time pressure during the event. The initial failed recovery attempt lasted for 33 seconds, from the airplane violently pitching down to the PIC declaring a PAN PAN. It took another 65 seconds to complete the second recovery attempt, the PM having restored flight control switches A and B at that point (AAIB, 2010b, p. 6).

7.4 Discussion of the Findings

Focus Question One: Why did it make sense for the crew to continue the test program?

The incident flight was commanded by one of the most experienced B737 check flight pilots in the airline community. A manual reversion check was a routine test point for him, including first-hand experience with recovering an airplane upset generated by incorrectly adjusted elevator tabs. He commanded the ferry flight prior to the maintenance event and provided verbal and written notes about the required sense and amount of elevator tab adjustments to his airline's technical representative. Unbeknown to him, a series of communication and coordination failures resulted in a grossly out-of-trim elevator system handed back to his crew for in-flight verification.

Automated flight control system functions compensated for and masked the out-of-trim condition up until test point entry. A series of interlinked adaptive failures between the various organisations involved in supporting the airline's operations led to a test configuration that closely resembled an experimental test flight. The crew members were neither aware, nor trained for such a scenario. Prior to test point entry, there were no warnings generated by the system that could have indicated to the pilots that the automated controller was at risk of decompensation.

Focus Question Two: Why did the recovery attempt work / fail to work?

In terms of recovery, this case study is unique, as it revealed an initial failed attempt, followed by an ad-hoc, then final recovery phase.

The failure scenario was not a complete surprise to the PIC, who correctly diagnosed the event and formulated a workable recovery plan. The initial attempt failed due to a growing communication and coordination gap between the PIC/PF and the PM. The ad-hoc phase involved local adaptations by an automated controller (Mach trim system) and the PIC/PF who was at risk of complete saturation due to cognitive and physical demands while attempting to regain control.

In this instance, while the primary controllers had different objectives, the local optimums achieved by the automated system and the human pilot reinforced each other and their combined control input was sufficient to overcome the aerodynamic loads on the elevator surfaces that ultimately led to a successful recovery. A full recovery could only be achieved, though, when the commander regained the capacity to scan his cockpit instruments and share his updated model with his less experienced copilot, bridging the communication and coordination gap between them.

The case study uncovered basic patterns in adaptive system failures between the main control units, including the a) the crew (PIC/PF and PM), b) the automated flight control systems, c) the airline’s management units, d) the two MRO units managing the layover, e) Boeing’s design and service engineering, and f) the regulatory oversight units (UK CAA and EASA).

As observed in the AAIB (2010b, p. 27) report, had no other factors assisted in the recovery, the commander would not have been able to arrest the descent. In this instance, the confluence of another set of systemic causal factors led to a fortunate scenario emerging where the automated system provided a silent but very powerful control input in support of the human operator.

Chapter 8 – CROSS-CASE ANALYSIS

8.1 Introduction

This study intended to investigate under what conditions, and through what causal paths, maintenance check flight accidents unfold. The purpose of the study could not be achieved by a traditional causal analysis method and a multiple case comparative analysis was performed. This chapter presents the results of the final data analysis phase that provided the response for the second and third research questions.

Research Question 2

Are there any common patterns in these occurrences? If there are common patterns, what do they reveal about the deeper structure of check flight accidents?

Research Question 3

Is there an alignment between underlying patterns and the risk control framework implemented by the airline industry?

As outlined in the methodology section, the Eisenhardt (1989, 1991, 2021) method served as a robust qualitative framework for the multiple case analysis part. Upon completing the initial case studies, the method requires searching for cross-case patterns that form the basis of a tentative grounded theory. Building a grounded theory from multiple cases is a highly iterative process (Eisenhardt, 2021; Yin, 2014). The following sections focus on major iteration loops that involved building the initial case group, extending the constructs with input from within-case analysis results, and confirming (or disconfirming) tentative patterns through replication.

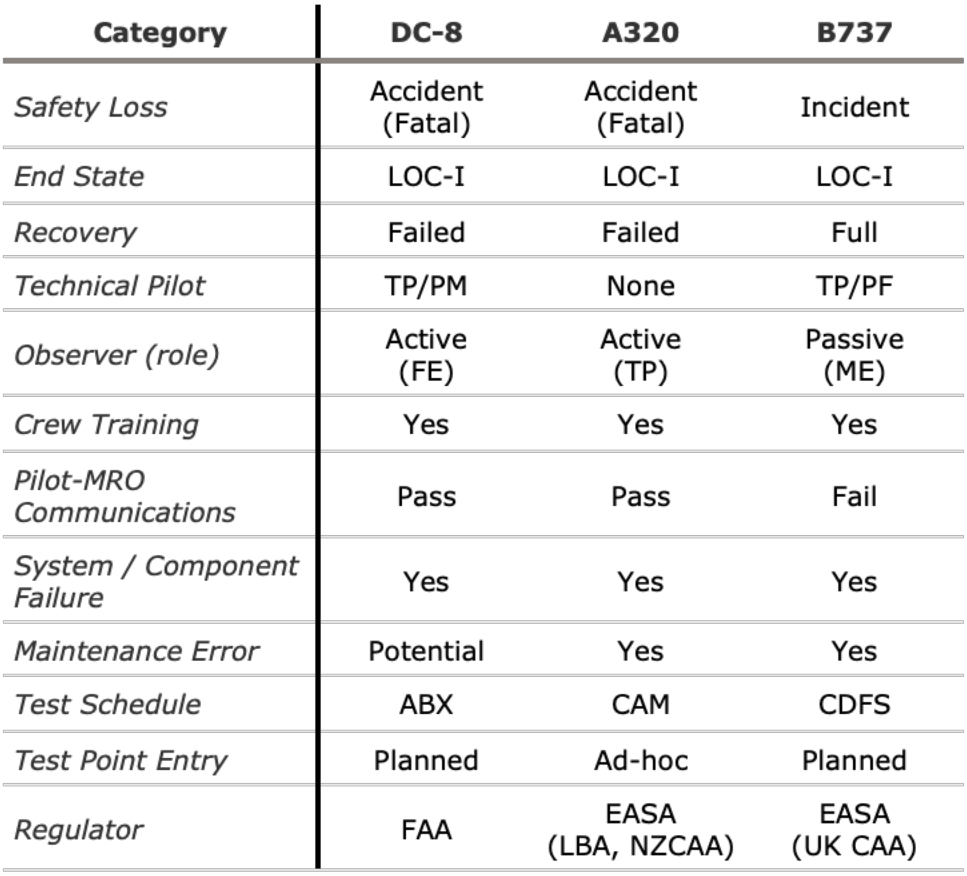
8.2 Initial Case Group

In accordance with the theoretical sampling criteria established in chapter three, three landmark events were selected from the overall population of 21 check flight investigation records. The three landmark events formed the initial case group for the multiple case study. Table 7 provides the key characteristics for each case in terms of the selected tentative categories.

As seen in Table 7, the cases represent different airliner generations and regulatory frameworks. Case A (DC-8) and Case B (A320) form a *matched pair* with similar antecedents and share the same outcome. When Case A (DC-8) and B (A320) are compared to Case C (B737), the pairs form *polar types* in terms of safety loss outcomes. There is variability across technical pilot skills and positions, active and passive observer roles, and known communication issues between maintenance engineers and the crew. While not visible in the summary, there is a different airline and maintenance organisation behind each case, as documented in the previous chapters.

Table 7

Initial Case Group and Categories



Category	DC-8	A320	B737
Safety Loss	Accident (Fatal)	Accident (Fatal)	Incident
End State	LOC-I	LOC-I	LOC-I
Recovery	Failed	Failed	Full
Technical Pilot	TP/PM	None	TP/PF
Observer (role)	Active (FE)	Active (TP)	Passive (ME)
Crew Training	Yes	Yes	Yes
Pilot-MRO Communications	Pass	Pass	Fail
System / Component Failure	Yes	Yes	Yes
Maintenance Error	Potential	Yes	Yes
Test Schedule	ABX	CAM	CDFS
Test Point Entry	Planned	Ad-hoc	Planned
Regulator	FAA	EASA (LBA, NZCAA)	EASA (UK CAA)

8.2.1 Pairwise Comparisons

Comparing Case A (DC-8) and Case B (A320). The two cases share a lot of common antecedent features. The key differences are the regulatory framework (FAA v. EASA), the way the in-flight test was conducted (planned v. ad-hoc), and the airliner generations involved (Generation 1 v. Generation 4). There was an authorized technical pilot on board in Case A, however, he acted in an instructor capacity and did not fly the DC-8 during the event. In this sense, both flying pilots in Case A and Case B had similar responsibilities and very low exposure to check flying.

Working with matched pairs that resulted in the same outcome, it is necessary to compare the processes involved that led to the fatal accidents. Despite the differences between generations, the regulatory systems, and in-flight test plan management, the safety loss was the same. When cross-checked with existing theory, it appears that, in isolation, better planning for in-flight test management does not offer improved safety outcomes for check flying.

Comparing Case A (DC-8) and Case C (B737). Case A (DC-8) and Case C (B737) are opposites in terms of the outcome variables. Both cases describe a loss of control scenario, but the DC-8 crew was unable to recover the airplane that resulted in multiple fatalities and a hull loss. In contrast, the experienced technical pilot at the controls of the B737 was able to arrest the airplane’s rapid descent. This fact was noted as a potential theme, subject to further validation. In contrast, when comparing communication breakdowns between the pilot and maintenance engineers, Case A and C do not support existing notions that providing the pilot with a rich context about maintenance action taken during the layover would achieve better outcomes.

Comparing Case B (A320) and Case C (B737). Again, Case B (A320) and Case C (B737) are opposites in terms of safety loss. When compared to the previous analysis, in this instance, the underlying regulatory framework was the same for both occurrences. During the A320 accident flight there was no technical pilot at the controls and there was no communication breakdown noted between the pilot and maintenance engineers. The notion about the potential safety benefits of having an experienced technical pilot at the controls holds for all three cases in the selected sample.

At the same time, Veillette's (2009) observation that providing the pilot with a rich context about maintenance action taken during the layover cannot be substantiated as a necessary precursor to better safety outcomes in airline operations. Subject to further analysis, it appears that modern airliners reached a level of complexity that potential in-flight failure modes experienced by the pilot (surprise events) may not be easily linked to discrete maintenance action taken. Case A and B provide examples where, prior to the accident flight, the maintenance crew was likely not even aware which maintenance action was to be seen as critical in hindsight.

Proposition One: For a highly automated airliner, without prior knowledge of the in-flight failure scenario, it is unlikely that maintenance *engineers can provide the pilot with a complete range of potential surprise events.*

8.2.2 Extended Categories

Following the initial tentative findings, the three case studies were revisited to gain a better insight about the role technical pilots and highly automated systems play in maintenance check flight outcomes. Data collected for the purposes of within-case analyses, specifically adaptive system failures, was the focus of this iteration.

As outlined in the previous three chapters, within-case analyses revealed basic patterns of adaptive system breakdowns in the selected accident investigation reports. It is important to highlight that the unit of analysis is not the same for a case study and the cross-case pattern analysis described in this chapter. By definition, a within-case analysis focuses on patterns at the level of adaptive units in a network, all within the bounds of a case. To identify categories suitable for a grounded theory, it is important to either retain the proposed category at the adaptive unit level, or to represent a group of adaptive units as a path which propagates the adaptive failure.

In this instance, case data suggested categories at both levels of abstraction. Table 8 presents a summary of within-case analysis results in the case studies. Basic adaptive traps are listed for each adaptive unit (rows) and case study (columns). After several iterations between case stories and the emerging accident theory, adaptive failures led to the following tentative constructs: a) *technical pilot decision making*, b) *control transfer* (from automated systems to the crew), and c) the significance of time represented in a broader *safety margin* category.

The remaining basic patterns in adaptive system failures were grouped into two main pathways. In-service breakdowns between airline, MRO, ATC, and regulatory units form an Operations pathway, shaping the crew's adaptive response. Design-induced adaptive breakdowns between the airframe manufacturer, MRO, airline, and regulatory units form another pathway, shaping the airplane's response.

Table 8

Adaptive System Failures

Adaptive Unit	**DC-8**		**A320**		**B737**	
Crew (PF, PM, Observer)	Communication	Coordination	Misalignment	Outdated plans	Residual Misalignment	Outdated plans
	Cross-purposes	Safety margins	Cross-purposes	Decompensation	Local adaptation	Saturation
Automated Systems	Silent transfer		Silent reconfiguration		Cross-purposes	
				Decompensation	Local adaptation	Decompensation
Maintenance Organization(s)			Communication	Coordination	Communication	Coordination
Air Traffic Control	Local adaptation		Local adaptation			
Airline Management	Outdated plans				Communication	Coordination
Design / Support	Potential surprise	Unreliable warning	Flight deck effects	Maintenance instructions	Elevator tab design	Maintenance instructions
Regulations	FAA oversight			EASA rules	CAA oversight	EASA rules
Time	120 sec to Recovery	82 sec to GPWS warning	62 sec post-Stall warning		33 sec Initial Recovery	65 sec Full Recovery

8.3 Replication

At this stage of the iterative process, preliminary relationships were emerging from the patterns observed in accident and incident case data. In line with the prescribed research method, it was necessary to validate and refine the emerging constructs and the underlying causal dynamics through a replication process. Two additional cases were selected for the purposes of confirming or disconfirming proposed constructs and categories: Case D (A340 accident at Salamanca in 2002), and Case E (E190 accident at Alverca in 2018).

8.3.1 Synopsis for Case D (A340)

During a post-maintenance check flight, the commander decided to test the don't sink warning of the GPWS system and requested ATC approval to conduct two flypasts at a low height. When on final approach, the aircraft started deviating to the left of the runway axis. The aircraft continued descending while the deviation was increasing. The main landing gears touched down on the apron and the left main wheel bogie collided with a recessed electrical channel cover. The right outboard flap collided with an 11 m high sentry box adjacent to the airport fence. The airplane suffered extensive damage, a 3m section of the flap was missing, the gears could not be retracted during climb, and a flap track fell onto the runway upon landing. Both main landing gear assemblies and the complete outboard flap installation had to be replaced. The Spanish authorities recorded the event as a serious incident (CIAIAC, 2003). This study recorded the 2002 Salamanca event as an accident, in line with ICAO (2020a) criteria (refer Section 1.8 Definitions).

8.3.2 Synopsis for Case E (E190)

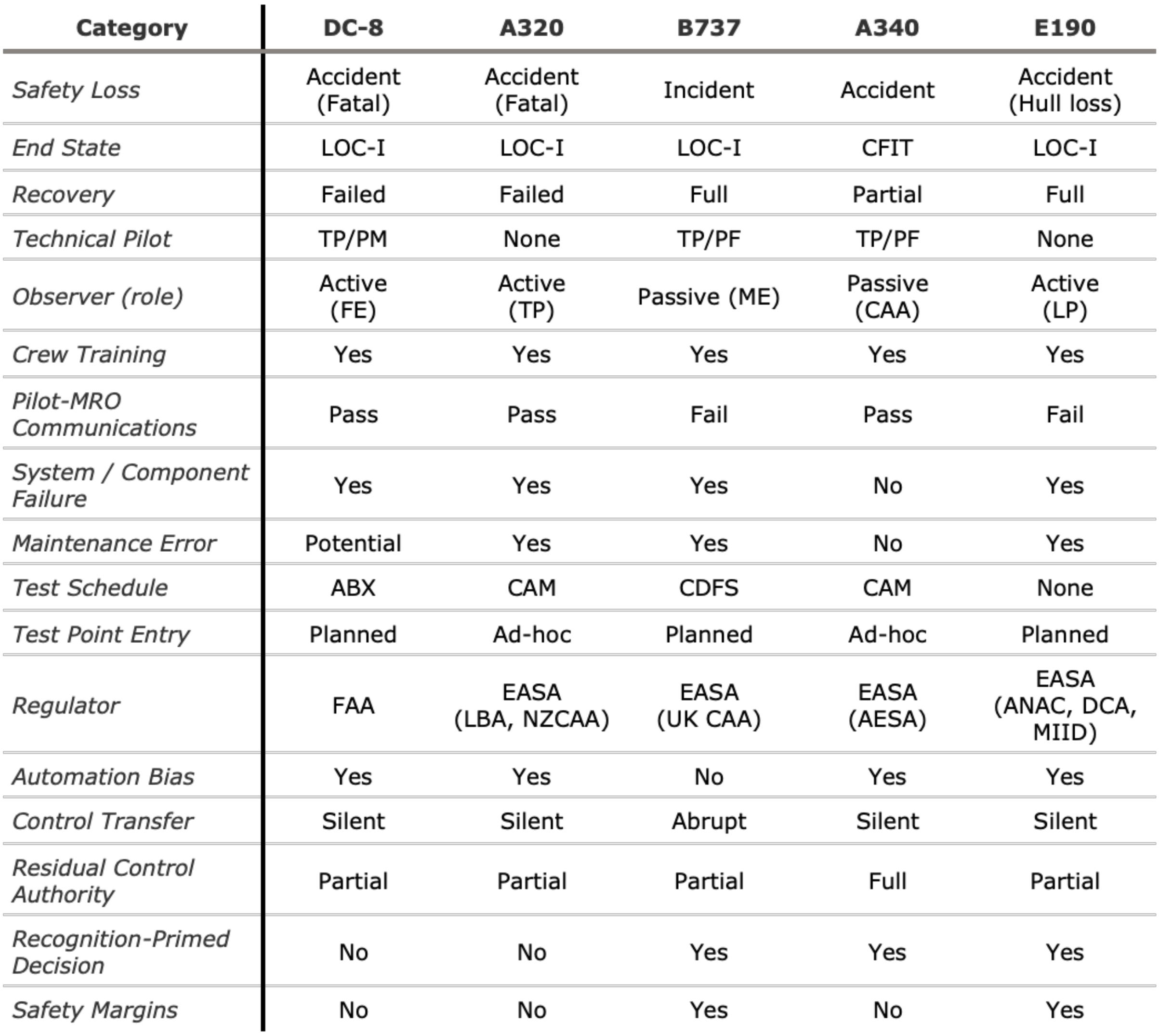
The airplane departed for a post-maintenance validation and ferry flight. Immediately after take-off, in adverse weather conditions, the crew realized that they had no effective control of the airplane and declared an emergency, while trying to diagnose the abnormal aircraft attitude. After two non-stabilized approaches, they managed to land on a third attempt. One crew member suffered minor injuries and the airplane was written off. The investigation found that the aileron cables were incorrectly installed during maintenance, which resulted in both ailerons acting in the opposite direction of control yoke commands (GPIAAF, 2020).

8.4 Extended Cases and Categories

A final iteration was performed to validate the proposed categories and themes across the five cases. Upon revisiting the preliminary relationships, further refinements were necessary to properly describe the emerging cross-case patterns. In lieu of the proposed technical pilot decision making category, *recognition-primed decisions* and *automation bias* were introduced to better reflect the patterns observed in case stories. Finally, *residual control authority* was added as a new category, refining the *control transfer* patterns revealed in the research notes. Table 9 provides an overview of extended cases and categories annotated.

Table 9

Extended Cases and Categories



Category	DC-8	A320	B737	A340	E190
Safety Loss	Accident (Fatal)	Accident (Fatal)	Incident	Accident	Accident (Hull loss)
End State	LOC-I	LOC-I	LOC-I	CFIT	LOC-I
Recovery	Failed	Failed	Full	Partial	Full
Technical Pilot	TP/PM	None	TP/PF	TP/PF	None
Observer (role)	Active (FE)	Active (TP)	Passive (ME)	Passive (CAA)	Active (LP)
Crew Training	Yes	Yes	Yes	Yes	Yes
Pilot-MRO Communications	Pass	Pass	Fail	Pass	Fail
System / Component Failure	Yes	Yes	Yes	No	Yes
Maintenance Error	Potential	Yes	Yes	No	Yes
Test Schedule	ABX	CAM	CDFS	CAM	None
Test Point Entry	Planned	Ad-hoc	Planned	Ad-hoc	Planned
Regulator	FAA	EASA (LBA, NZCAA)	EASA (UK CAA)	EASA (AESA)	EASA (ANAC, DCA, MIID)
Automation Bias	Yes	Yes	No	Yes	Yes
Control Transfer	Silent	Silent	Abrupt	Silent	Silent
Residual Control Authority	Partial	Partial	Partial	Full	Partial
Recognition-Primed Decision	No	No	Yes	Yes	Yes
Safety Margins	No	No	Yes	No	Yes

8.4.1 Competent Technical Pilot at the Controls

The study found strong evidence in support of the notion that better safety outcomes are achieved when technical pilots fly the airplane during critical test points. For example, Case C (B737) describes a successful recovery sequence after an abrupt LOC-I scenario. The case not only demonstrates the pilot's flying skills. The earlier decision to enter the manual reversion test point with a sufficient altitude margin was equally important. In contrast, Case D (A340) provides an example where safety margins were eroded to the point that a full recovery was not possible. Again, the experienced technical pilot's reaction and flying skills were essential in limiting the damage suffered by the airframe during the CFIT encounter.

Proposition Two: When a competent technical pilot flies the airplane during a critical test point, there is an increased likelihood of a successful recovery attempt.

As Table 9 indicates, all cases confirm the above proposition. Case A (DC-8) involved two management pilots with minimal check flying experience. Both were experienced in routine flight operations, and the PM was formally authorized to conduct post-maintenance test flights, but the PF had only started his DC-8 check flying the day before the accident flight. None of the GXL crew members involved in Case B (A320) had ever received any form of training in non-routine operations. The observer had a formal technical pilot authorization from ANZ but he had no relevant experience in check flying the A320. He attended an unsupervised flight simulator session to familiarize himself with some of the planned test points (BEA, 2010, p. 75). Both accidents resulted in multiple fatalities.

Finally, Case E provides an example where a reinforced crew of three line pilots were tasked to conduct a combined validation and ferry flight. None of the crew members had any recorded check flying experience prior to the accident flight. After a couple of failed attempts, the experienced airline captain (PF) managed to regain partial control of the airplane, but not without overstressing the airframe during the LOC-I event, that resulted in a complete hull loss (GPIAAF, 2020, pp. 102-103).

8.4.2 Maintenance Error and System Failures

As Case D (A340) illustrates, during a check flight program, a surprise event can develop into a full-scale accident without experiencing any system or component failures or any form of maintenance-induced error. In this instance, the ad-hoc decision taken by an experienced technical pilot to test a warning system function beyond design specifications was a critical precursor to the unfolding event (CIAIAC, 2003, pp. 33-34). All other cases in the comparative analysis annotated a maintenance error, a system or component failure, or both.

Proposition Three: Maintenance error(s) and System/Component Failure(s) are not a necessary causal condition for maintenance check flight occurrences.

The earlier observation (Proposition One) that maintenance engineers are not necessarily able to brief the pilot on a complete range of potential in-flight surprise events is closely related to this latest theme. While the primary objective of in-flight verification remains to mitigate the risk of inducing maintenance errors to safety critical systems, it does not follow that novel or unexpected failure scenarios that are not related to disturbed systems cannot be experienced by the crew.

8.4.3 Automation Bias and Control Transfer

Woods (2018) highlighted the fundamental problems associated with design concepts that switch control from automated systems to a human operator upon reaching preset control authority limits. Without exception, all four airliner generations studied here follow that typical adaptive pattern between automated systems and the pilots. Similar to routine line operations, this design philosophy can lead to unexpected breakdowns, especially when automated control is transferred in a silent fashion (Woods & Sarter, 2000).

Case C (B737) is the only example in the extended study that did not involve a silent control transfer. To the contrary, the PF experienced immediate mechanical feedback, in the correct sense, through the control column when hydraulic assistance was removed from the grossly out-of-trim elevator control system. Only minutes earlier, when testing the same control system with single hydraulic boost, the automation had successfully masked the actual trim condition. The fact that the incident PF was ready for an abrupt control transfer, having been confronted by a similar aircraft upset scenario in the past, indicates a potential theme in navigating recovery attempts.

All other cases corroborate the underlying theme. The DC-8 crew in Case A was likely influenced by expectation bias, awaiting a smooth airplane stall response, conditioned by incorrect and misleading flight simulator characteristics. The GXL crew in Case B (A320) experienced a startle event when a decompensating FBW system transferred control to them in direct law, further degrading to abnormal attitudes logic during the event (BEA, 2010; Martin, 2014). In Case D, the A340 EGPWS system logic ensured that the responsibility for managing ground clearance had transferred to the PF while in the circuit, long before he attempted the approach to the runway (CIAIAC, 2003). And finally, Case E (E190) is another example when the crew was surprised (and potentially startled) upon discovering that a significantly degraded flight control system was transferred to the PF in line with the built-in system logic and architecture (GPIAAF, 2020).

Proposition Four: When automated systems transfer control to the human pilot in a silent fashion, especially when the pilot is influenced by automation bias, the recovery attempt is likely to fail, unless safety margins enable repeated recovery attempts.

8.4.4 Recognition-Primed Decision Making

Inverting the logic in the previous section, the case studies were revisited to find out what conditions were supportive of recovery attempts. Klein et al. (1993) RPD model was a good fit to the data, not only matching the decision making environment, but also revealing common themes in investigation reports and post-incident interviews. The author's personal insight echoes the finding that recognition plays a crucial role in resolving unexpected check flight scenarios. The cross-case data did not contain cognitive traces for analysis, but limited inferences can be made from temporal events and CVR transcripts (Klein & Wright, 2016).

Cases C (B737), D (A340), and E (E190) provide evidence of critical recognition-primed decisions made prior to successful recovery attempts. The decision context was different in each case, primarily shaped by available safety margins (time / altitude) and the remaining control authority of the pilot, both in a physical and a cognitive sense. The B737 technical pilot instantly recognised the failure condition and formulated a suitable recovery plan but did not have sufficient residual control authority to implement the plan on his own. Similarly, the A340 technical pilot instantly recognised the collision risk but did not have sufficient time and remaining obstacle clearance to avoid the collision. Finally, the E190 pilots took much longer to recognise that their aileron controls were reversed, however, through failed recovery attempts, they formulated a workable plan to recover the airplane with a significantly reduced control authority.

Case A (DC-8) documents a trace where both pilots recognised the initial stall scenario. At the same time, it is unlikely that they recognised the fact that the airplane entered a deep stall. They did not have sufficient remaining control authority and safety margins to formulate and implement a successful recovery plan. Finally, the A320 crew involved in Case B had some residual control authority but were confronted by a novel challenge that they could not resolve within their diminished safety margins at a low height above ground. The third crew member (ANZ technical pilot) incorrectly diagnosed the airplane upset as a response by flight envelope protections, likely influenced by automation bias.

Proposition Five: When a competent technical pilot recognises the in-flight failure scenario and has sufficient residual control authority and safety margins when formulating a response, the recovery attempt is likely to be successful.

8.4.5 Design and Operations Pathways

Table 9 revealed additional basic patterns in systemic failures between interlinked adaptive units. The cross-case comparison uncovered a considerable level of variety in underlying dynamic relationships (causal links) between the adaptive units. For the purposes of theory building, two main pathways were formed from case stories: a) breakdowns between adaptive units in *Operations*, and b) breakdowns induced by design or certification issues through a *Design* pathway.

Within-case analyses embedded in chapters four, five, and six provide ample evidence for both pathways, and as such, the relevant sections are not repeated here. Following the basic patterns observed in each case, adaptive system failures manifest in automated systems and the pilot pushed to the limits of their adaptive capacity, working at cross-purposes, or the active controller working to outdated models and plans.

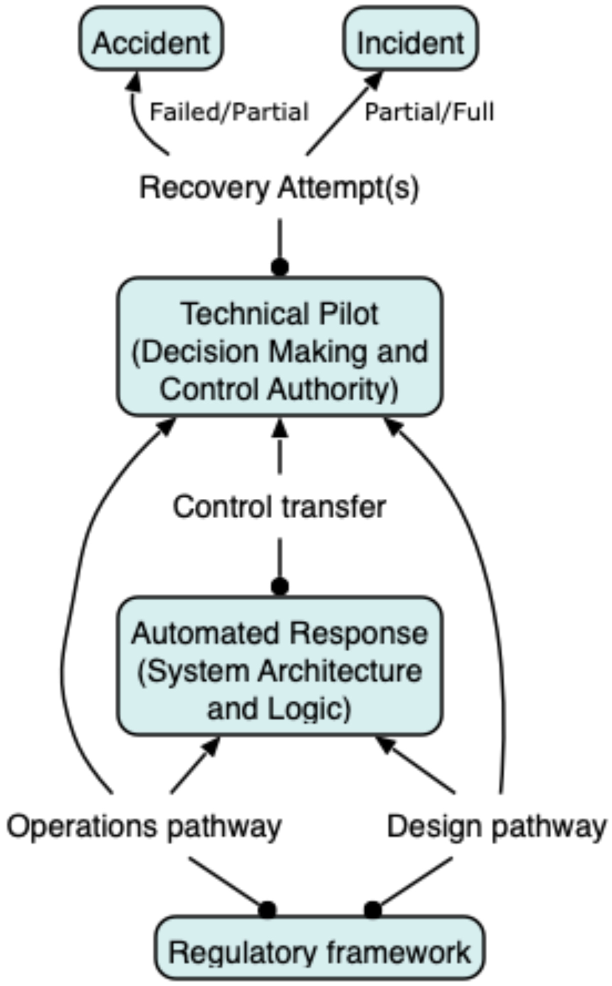
Proposition Six: Interlinked adaptive breakdowns can form design and operational pathways that diminish the automated systems' or the pilot's ability to respond to unexpected or novel in-flight scenarios.

8.5 New Grounded Theory

The objective of the research study was to investigate under what conditions, and through what causal paths, maintenance check flight accidents unfold. More specifically, the cross-case study was testing the second and third research questions. The overall result of the cross-case comparison was a new accident model for maintenance check flight operations in an airline environment, illustrated in Figure 40.

Figure 40

Accident Model for Airline Maintenance Check Flights



8.5.1 Common Patterns

Research Question 2

Are there any common patterns in these occurrences? If there are common patterns, what do they reveal about the deeper structure of check flight accidents?

The central organising theme in the accident model is a common design principle across all airliner generations that requires automated systems to transfer control to the human pilot at preset design limits. In system safety terms, the airplane’s behaviour during a check flight event is deterministic, defined by the logic and constraints embedded by the designer in the system’s architecture and the rules given to the automation about when and how to transfer control to the human pilot.

When control is transferred to the pilot, a recovery attempt is more likely to succeed if the pilot recognises the in-flight failure scenario and has adequate control authority and remaining safety margins (time / altitude) when formulating a response (Proposition Five). When entering critical test points, it is preferred to have a competent technical pilot at the controls, as relevant training, skills, and experience, can increase the likelihood of a successful recovery attempt (Proposition Two).

Check flight occurrences show patterns in basic adaptive system failures. These breakdowns form dual design and operational pathways that are interlinked and diminish the automated systems' and the pilot’s ability to respond to unexpected or novel in-flight scenarios (Proposition Six). For a highly automated airliner, neither the designer, nor maintenance engineers can provide the pilot with a complete range of potential surprise events (Proposition One).

8.5.2 Alignment with Existing Safety Response

Research Question 3

Is there an alignment between underlying patterns and the risk control framework implemented by the airline industry?

As highlighted in the literature review chapter, the airline industry adopted a risk control framework that focuses on preflight planning and preparations and a heavy emphasis on technical pilot training (Airbus, 2015; FSF, 2011a). Previous studies on the topic assume that maintenance-induced errors and systems or components disturbed during maintenance are the primary concern for check flight operations (Veillette, 2009).

On a positive note, the study found cross-case evidence in support of the emphasis placed on technical pilot training and better overall competency levels (refer Proposition Two) (Airbus, 2015; Poprawa, 2015). Better test planning and preparations have undeniable safety benefits, especially in terms of allocating and maintaining adequate safety margins during check flight operations (Propositions Four and Five). At the same time, the study uncovered a fundamental misalignment between the plan and prepare approach and the underlying causal patterns in the accident model. Maintenance check flight occurrences involve systemic failures that propagate through operational and design pathways, and as such, the safety risk cannot be mitigated by airline and maintenance operations alone (Proposition Six).

The main objective of maintenance check flights is to safeguard revenue flights from the risk of inducing maintenance errors to safety critical systems. It needs to be highlighted, though, that maintenance error(s) and system or component failure(s) are not a necessary causal condition for check flight accidents (Proposition Three). During check flight operations, the crew may experience a range of novel or unexpected scenarios that are not related to systems disturbed during the maintenance event.

CHAPTER 9 – SUMMARY, DISCUSSION, AND CONCLUSIONS

9.1 Introduction

In the preceding chapters, case study data and cross-case analysis results have been presented. This chapter consists of a summary of the overall research study, discussion of findings, implications for practice, recommended further research topics, and concluding statements. The purpose of this chapter is to revisit the original aims of the study, review the key constructs and findings in more detail, and to emphasize the practical implications of the new accident model.

9.6 Summary of the Study

This section briefly restates the purpose and structure of the research study, followed by major findings related to the causal theory of maintenance check flight accidents. The summary places an emphasis on the research findings that will be discussed in even more detail in the next section.

The purpose of the study was to develop causal explanations for maintenance test flight accidents. Compared to the sustained research interest in improving airline safety, non-routine flights are rarely mentioned in academic literature, despite the considerable safety losses suffered over the years. This study aimed to address that gap by revisiting multiple landmark accidents through a robust qualitative research framework.

Accident investigation reports published by relevant government safety agencies provided the necessary research data. A systematic review of selected accident investigation databases identified 101 non-routine flight investigation accidents and incidents suffered by Western-built airliners since June 1988. The 101 non-routine investigations included 22 post-maintenance check flight occurrences across four airliner generations. The number of check flight accidents was validated against independent safety statistics available in grey literature (Boeing, 2022).

The next step of the qualitative analysis involved an empirical study that revealed 45 causes and 54 contributing factors that were annotated in check flight investigation reports. From these first-order constructs, the study elicited nine common categories, ranging from regulatory oversight, through operational and design issues, to flight crew decision making. A knowledge map built from the output of the empirical study informed the final data analysis step, and later served as independent means when validating the results.

The final part in the research study followed a multiple case grounded theory method (Eisenhardt, 1989, 2021). Three landmark investigation reports were selected for the initial case studies, each case representing a different airliner generation. The cross-case analysis revealed 5 main themes and 6 tentative propositions that were successfully replicated across an extended case sample. Finally, the results were assembled into a single accident causation model that reflects the main causal dynamics, as explained in the next section.

The study included three research questions:

Research Question 1

What causes and contributing factors are annotated in check flight investigation reports?

Research Question 2

Are there any common patterns in these occurrences? If there are common patterns, what do they reveal about the deeper structure of check flight accidents?

Research Question 3

Is there an alignment between underlying patterns and the risk control framework implemented by the airline industry?

The research questions were answered qualitatively. Question 1 was tested using the first-order constructs and second-order categories identified in accident and incident investigation reports. To answer Question 2, the cross-case analysis results were combined with the basic adaptive patterns revealed in individual case studies. The new accident model addressed the second part of Question 2 when describing the deeper structure of check flight accidents. Finally, Question 3 was addressed by engaging existing safety literature during the cross-case analysis.

9.3 Discussion of the Findings

The overarching goal of the study was to discover under what conditions, and through what patterns, maintenance test flight accidents unfold. This section discusses how the results support or contradict the existing knowledge base for each research question.

Research Question 1

What causes and contributing factors are annotated in check flight investigation reports?

Previous studies on post-maintenance test flight safety were either based on very limited samples, or a sample that mixed various aircraft types and operations (Poprawa, 2015; Veillette, 2009). The first research question was borne out of that significant literature gap and the need to establish a baseline for Western-built commercial jet airplanes. The study identified 22 check flight occurrences within scope. It is noted that one of the historical investigation records could not be located. From the remaining 21 reports, the initial data analysis revealed nine causal categories. The most frequently cited causes were flight crew error, maintenance error, airworthiness management problems, and design issues, closely followed by system or component failures. The remaining causal factors fell into the crew training, check flight regulations, environmental, or undetermined categories.

The categories were closely aligned to the results of a similar independent study that evaluated probable and contributory causes of major aviation accidents in the United States between 1996 and 2003, as adopted by the NTSB (Holloway & Johnson, 2004). Other than the 1996 DC-8 accident at Narrows, the NTSB study did not include non-routine occurrences. Despite ~~of~~ the different sample, the close correlation between the categories was not surprising. As a Contracting State, the United States follows ICAO SARPs that define the format and content of standard investigation reports, including example wording for findings and a common taxonomy for accident and incident reporting purposes (ICAO, 2020a).

In contrast, the results did not align with a previous study on post-maintenance test flight safety. Veillette (2009) evaluated 128 incident records, a mix of voluntary safety reports and NTSB investigations across fixed-wing and rotary aircraft types. He selected categories that emphasised the check pilot’s perspective, citing the extra workload and abnormal procedures, abnormal crew coordination, system or component failures, and even distractions during preflight preparations as major threats.

Poprawa’s (2015) contribution on the topic relied on a randomly selected sample of maintenance test flight accidents suffered by Eastern- and Western-built commercial jet airplanes. In lieu of a causal analysis, he opted for a different approach when highlighting that LOC-I remains the dominant occurrence category during check flights. He found that regulatory oversight, crew training, experience, qualifications, and the ability to recover from complex LOC-I events remain unresolved open risk elements. These constructs correlate to the empirical results in this study.

Research Question Two

Are there any common patterns in these occurrences? If there are common patterns, what do they reveal about the deeper structure of check flight accidents?

The research study identified the same fundamental problems at the heart of check flight accidents. All four airliner generations follow a similar design principle that requires automated systems to transfer control to the human pilot as a last resort. As a result, check flight occurrences reveal similar patterns in adaptive system breakdowns, reflected in five common themes:

1. Competent technical pilot at the controls: When automation reaches its capacity to handle an unexpected or novel scenario (surprise event) during a critical test point, the human pilot needs to have the right training, skills, and experience to be able to take control and recover from the undesired airplane state. If the pilot flying is a competent technical pilot, the study found that there is an increased likelihood of a successful recovery attempt (refer Proposition Two in the previous chapter).
2. Maintenance error and system or component failures: While the primary objective of check flying remains to treat the risk of maintenance-induced errors degrading flight safety critical systems, it remains a common misconception that either maintenance error or system or component failures are a necessary causal trigger for these occurrences (Proposition Three). Modern airliners have reached a level of complexity that maintenance engineers cannot provide the crew with a full range of potential surprise events that may be linked to systems disturbed on the ground. And in line with the quoted design principle, automated systems may be certified by simply assuming that any novel or unexpected scenarios will be handled by the pilot (Proposition One).
3. Automation bias and control transfer: Building on the previous theme, when automated systems ultimately reach their limit and

transfer control to the pilot, the pilot needs to know that a transfer took place. If control is transferred in a silent fashion, especially when the pilot is influenced by automation bias, there may not be sufficient time remaining to recognise the situation and initiate a recovery attempt (Proposition Four).

4. Recognition-based decision patterns: When the previous theme was validated through case histories and available literature, a new theme emerged in a decision making context. Incident traces revealed that before initiating successful recovery attempts, technical pilots recognised the in-flight failure scenario and were able to form a response when they had enough time and sufficient residual control authority to do so (Proposition Five).
5. Design and Operational pathways: Finally, the cross-case analysis also revealed a common theme in the way systemic causal factors degrade the system's capacity to respond to unexpected or novel in-flight scenarios. The study found that systemic breakdowns form design and operational pathways that put the pilot and automated systems at risk of saturation, or in a worst case scenario, force one or both primary controllers to decompensate (Proposition Six).

The overall result of the analysis was a new accident model of airline post-maintenance check flight causation, illustrated in Figure 40 (see previous chapter). The model reflects the common themes outlined in the previous paragraphs. Due to the considerable variability in underlying causal dynamics between controllers (adaptive units), the model focuses on the critical causal relationships between the crew, automated systems, and the regulatory framework, linked by operations and design pathways. The new accident model emphasises the decisive role played by the interactions between the pilots and highly automated systems. It is virtually impossible to eliminate all the potential systemic causal factors that may lead to adaptive breakdowns and failures. Irrespective of the propagation pathway followed by those failures, the risk of check flight accidents can only be treated if the cockpit interface and the system design support the pilot in recognizing the in-flight scenario and provide sufficient warning and time before transferring responsibility for initiating a recovery attempt.

Research Question Three

Is there an alignment between underlying patterns and the risk control framework implemented by the airline industry?

The study uncovered a misalignment between basic patterns of adaptive breakdowns involved in check flight accidents and the safety management framework that was supposed to mitigate the risk of check flight accidents. The plan and prepare approach has a singular focus on the operations pathway and technical pilot competency, and disregards the risk posed by known safety hazards that can propagate through the alternative design pathway. In terms of common themes uncovered in this study, the current risk management framework only targets theme one (competent technical pilot), two (maintenance error), and five (operational pathway, in part).

Common theme three identifies the risk associated with automation bias and control transfer. Should automated systems reach their adaptive capacity limit and transfer control to the pilot during a check flight, the pilot needs to know that a transfer took place and there must be sufficient safety margins (time / altitude) remaining, so that the pilot can recognise the situation and initiate a recovery attempt.

Common theme four refers to the decision making context and RPD patterns, closely related to the previous theme. Successful recovery attempts rely on the technical pilot's ability to recognise the in-flight failure scenario and to formulate a response, provided the pilot has enough time / altitude remaining and sufficient residual control authority to do so. Unexpected or novel scenarios pose an unmitigated safety risk in the plan and prepare safety framework.

9.4 Implications for Practice

As highlighted earlier, the plan and prepare approach is not aligned with common accident patterns revealed by the research study. Check flight events cannot be isolated to the operational airworthiness domain. The recently adopted EASA framework is a step in the right direction, but additional risk controls need to be implemented to address the residual system safety risk.

Modern airliners introduced novel design and certification challenges and a range of unknown or unforeseen systemic interactions. Check flights remain an effective risk mitigation tool for those novel systemic hazards, however, will continue to rely on the human pilot as the ultimate fallback for all unexpected failure conditions. Neither the designer nor the airline operator can plan for unknown or unforeseen scenarios. That is the fallacy of the plan and prepare approach adopted by the industry.

In lieu of blaming the pilot when automated systems cannot cope with the unfolding scenario, the industry needs to reinforce recovery margins. On the operational side, sufficient time and altitude brackets can only be achieved by planning every test point for a worst-case scenario. And on the designer’s side, the information gap between automated systems and human operators (ground and flight crew) must be revisited to support check flight mission objectives, which are fundamentally different from routine commercial operations.

The current safety paradigm demands that accident investigations describe a plausible sequence of events. That demand can efficiently be met by relatively simple RCA techniques, linear causal chains, nominating probable causes, and the like. This simple mechanical view of the world, however, does not apply to complex systems, like highly automated airliners. Academia developed a range of more sophisticated accident models, which more accurately describe what happens when modern systems fail. Are those new models a better way to approach system safety? Not necessarily, as system safety tools need to be practical, first and foremost. If the model, analysis results and safety recommendations are only understood by a select group of academics, there is not much benefit to the practitioner.

This thesis aimed at finding the middle-ground by advocating for an approach which improves the fidelity of the accident model and, at the same time, remains practical by using relatively simple concept mapping tools. Other than being practical, the novel approach applied in this study offers an added benefit in accident prevention: common patterns can more easily be translated to long-term and robust positive safety outcomes, as the recommendations are not limited to individual traces of single events. And that is a much-needed improvement from the traditional fly-fix-fly approach to air safety investigations.

9.5 Recommendations for Further Research

The research study identified a safety critical area of unmitigated systemic risk associated with potential causal propagation through the design pathway. The issue is further aggravated by common patterns uncovered in human-machine interactions, influenced by automation bias and suboptimal control transfer. The findings from this research project can serve as the basis for further research to better understand what reasonably practical engineering controls are available to mitigate the residual system safety risk in the short- and medium-term.

The study also confirmed that check flight accident rates are relatively high compared to other non-routine operations. When compared to ultra-safe line operations, all forms of non-routine flights experience a very high rate of accidents and serious incidents. A similar research study that revisits investigation findings about positioning, ferry, and training flight occurrences may offer additional insights into the non-routine safety problem.

9.6 Conclusions

As stated in the introduction, the airline industry is justifiably proud of the ultra-safe operational record achieved by modern airliner generations during routine commercial operations. Unbeknown to the public, non-routine flights, such as ferry, training, or post-maintenance check flights, contribute to that excellent safety record by offering airline operations and engineering departments the means to shift higher risk operations to non-revenue sectors, when only essential crew members are on board.

This research project set out to better understand and characterize the risk involved in post-maintenance test flights and revealed the following major findings:

The study confirmed that check flight occurrences are not limited to older and less reliable aircraft generations. To the contrary, the latest FBW generation appears to have introduced new error traps and challenges during maintenance and flight operations, reflected in more serious safety losses over the years.

In response to the safety challenge, the airline industry adopted a new operational risk management framework, in line with a plan and prepare mindset recommended by airframe manufacturers. The research found a misalignment between the safety response and the underlying common causal patterns elicited from check flight investigation reports.

Complex systems introduced novel design and certification challenges. Those same challenges pose an elevated risk of introducing unexpected or unforeseen causal scenarios during maintenance or flight operations. As the mission objectives of routine and check flight sectors are not compatible, any potential breakdown between automated systems and the human pilot carries a higher risk of eroding remaining safety margins.

Automation is a rules-based, deterministic, pre-programmed response to external stimuli from the environment. When no acceptable integrity check is available during maintenance, we ask the human pilot to conduct a functional check on automated system(s) in-flight. The irony is that when the automated system malfunctions, for whatever reason, investigations tend to blame the human pilot as the probable cause for the incident or accident.

Automation has limited authorisation and capacity to compensate during in-flight scenarios. The crew intentionally challenges the limits of automation when conducting test points, be it an operational or functional check flight. When automated systems transfer control to the pilot, should the transfer be sudden or unexpected, the crew may not be ready to take control or there may be no time left for them to recover.

In summary, while post-maintenance check flights remain a very effective tool for protecting ultra-safe routine airline operations, the airline industry needs to adapt its safety response when it comes to mitigating the elevated systemic risk of experiencing a check flight accident or serious incident.

REFERENCES

Abbott, K. (2000). Human factors engineering and flight deck design. In C. Spitzer (Ed.) *Digital avionics handbook*. CRC press.

Abbott, K., Slotte, S., Stimson, D. (1996). *The interfaces between flightcrews and modern flight deck systems*. Federal Aviation Administration.

Abbott, T. S., & Rogers, W. H. (1993). Functional categories for human-centered flight deck design. In *[1993 Proceedings] AIAA/IEEE Digital Avionics Systems Conference* (pp. 66–74). https://doi.org/10.1109/DASC.1993.283569

Abzug, M. J., & Larrabee, E. E. (2002). *Airplane stability and control. Cambridge Aerospace Series*. Cambridge University Press.

Adler, J. E. (2008). Presupposition, Attention, and Why-Questions. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 748–764). Cambridge University Press.

Aeronautica Civil. (1996). *Aircraft accident report, Controlled flight into terrain American Airlines Flight 965 Boeing 757-223, N651AA near Cali, Colombia December 20, 1995*. https://www.fss.aero/accident-reports/dvdfiles/CO/1995-12-20-CO.pdf

Air Accidents Investigation Branch. (n.d.). *Air Accidents Investigation Branch reports*. https://www.gov.uk/aaib-reports

Air Accidents Investigation Branch. (1994). *AAIB Bulletin: 8/94 Incident Concorde Type 1 Variant 102, G-BOAA London Heathrow Airport 29 April 1994 at 2100 hrs.* (Report No. 8/94).

Air Accidents Investigation Branch. (1998). *Report on the incident to Boeing 737-236 Advanced, G-BGJI 15 nm north-west of Bournemouth International Airport on the 22 October 1995.* (Report No. 1/1998).

Air Accidents Investigation Branch. (2009). *AAIB Bulletin: 6/2009 Incident Boeing 757-204, G-BYAO Over North Sea/London Stansted Airport, Essex 22 October 2006 at 0835 hrs* (Report No. AAIB 6/2009).

Air Accidents Investigation Branch. (2010a). *AAIB Bulletin: 12/2010 Incident Falcon 2000, CS-DFE Biggin Hill Airport, Kent 11 November 2009 at 1259 hrs*. (Report No. 12/2010).

Air Accidents Investigation Branch. (2010b). *AAIB Bulletin: 9/2010 Serious Incident Boeing 737-73V, G-EZJK West of Norwich, Norfolk 12 January 2009 at 1545 hrs*. (Report No. 9/2010).

Air Accidents Investigation Branch. (2013). *AAIB Bulletin: 2/2013 Incident Boeing 757-2K2, G-LSAN Over the North Sea 7 August 2012 at 1535 hrs.* (Report No. 2/2013).

Air Accidents Investigation Branch. (2022). *AAIB Bulletin: 1/2022 Serious Incident Airbus A319-111, G-EZBD 13 July 2021 at 1048 hrs London Luton Airport*. (Report No. 1/2022).

Air Accident Investigation Unit. (1998). *Final Report AAIU 1998/011 MD-82 HB-INW London FIR 18th December 1997 Shannon FIR 19th December 1997.* (Report No. AAIU 1998/011).

Air Accident Investigation Unit. (2004). *Final Report AAIU 2004-004 UPS Boeing B747-200 Dublin Airport 12 May 2000.* (Report No. AAIU 2004-004).

Airbus. (2015). Functional check flights. *Safety first*. https://safetyfirst.airbus.com/category/magazine/

Airbus. (2022). *A statistical analysis of commercial aviation accidents 1958-2021*. Airbus S.A.S.

Aircraft and Railway Accidents Investigation Commission. (1996). *Aircraft accident investigation report China Airlines Airbus Industrie A300B4-622R, B1816 Nagoya Airport April 26, 1994*. Ministry of Transport. https://www.mlit.go.jp/jtsb/eng-air_report/B1816.pdf

Alexander, C. (1977). *A pattern language: Towns, buildings, construction*. Oxford University Press.

Allianz Global Corporate & Specialty. (2014). *Global aviation safety study – A review of 60 years of improvement in aviation safety*. Allianz Global Corporate & Specialty SE. https://www.agcs.allianz.com/content/dam/onemarketing/agcs/agcs/reports/AGCS-Global-Aviation-Safety-2014-report.pdf

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science*, *37*(2), 109–126. https://doi.org/10.1016/S0925-7535(00)00045-X

Aspers, P. (2009). Empirical Phenomenology: A Qualitative Research Approach (The Cologne Seminars), *Indo-Pacific Journal of Phenomenology, 9*(2)*,* 1-12*.* https://doi.org/10.1080/20797222.2009.11433992

Australian Transport Safety Bureau. (n.d.). *National Aviation Occurrence Database*. https://www.atsb.gov.au/avdata/

Aviation Herald. (n.d.). *The Aviation Herald Incidents and Accidents*. http://avherald.com/

Axelrod, R. (Ed.). (2015). *Structure of decision: The cognitive maps of political elites*. Princeton University Press.

Bainbridge, L. (1983). Ironies of automation. In G. Johannsen & J. E. Rijnsdorp (Eds.), *Analysis, Design and Evaluation of Man* (pp. 129–135). Pergamon. https://doi.org/10.1016/B978-0-08-029348-6.50026-9

Baker, S. P., Qiang, Y., Rebok, G. W., & Li, G. (2008). Pilot Error in Air Carrier Mishaps: Longitudinal Trends Among 558 Reports, 1983-2002. *Aviation, Space, and Environmental Medicine*, *79*(1), 2–6. https://doi.org/10.3357/ASEM.2200.2008

Benner, L. (1975). Accident investigations: Multilinear events sequencing methods. *Journal of safety research*, *7*(2), 67-73.

Bertalanffy, L. von. (1969). *General system theory: Foundations, development, applications*. George Braziller.

Billings, Charles E. (1991, August 1). *Human-centered aircraft automation: A concept and guidelines*. (NASA Technical Memorandum TM-103885). https://ntrs.nasa.gov/citations/19910022821

Billings, Charles E. (1996, February 1). *Human-centered aviation automation: Principles and guidelines*. (NASA Technical Memorandum TM-110381). https://ntrs.nasa.gov/citations/19960016374

Boeing. (2001). *Statistical summary of commercial jet airplane accidents Worldwide operations 1959-2000*. Boeing Commercial Airplanes.

Boeing. (2011). *Statistical summary of commercial jet airplane accidents Worldwide operations 1959-2010*. Boeing Commercial Airplanes.

Boeing. (2012). *Statistical summary of commercial jet airplane accidents Worldwide operations 1959-2011*. Boeing Commercial Airplanes.

Boeing. (2021). *Statistical summary of commercial jet airplane accidents Worldwide operations 1959-2020*. Boeing Commercial Airplanes.

Boeing. (2022). *Statistical summary of commercial jet airplane accidents Worldwide operations 1959-2021*. Boeing Commercial Airplanes.

Bradley, E. A. (1995). Determination of human error patterns: The use of published results of official enquiries into system failures. *Quality and Reliability Engineering International*, *11*(6), 411–427. https://doi.org/10.1002/qre.4680110605

Braune, R. J., & Graeber, R. C. (1992). Human-centered designs in commercial transport aircraft. *Proceedings of the Human Factors Society Annual Meeting*, *36*(15), 1118–1122. https://doi.org/10.1518/107118192786749702

Brennan, S. E. (1998). The grounding problem in conversations with and through computers. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication* (pp. 201-225). Hillsdale, NJ: Lawrence Erlbaum. http://www.psychology.sunysb.edu/sbrennan-/papers/brenfuss.pdf

Briere, D., Favre, C., & Traverse, P. (2000). Electrical flight controls, from Airbus A320/330/340 to future military transport aircraft: A family of fault-tolerant systems. In C. R. Spitzer (Ed.), *Digital avionics handbook*. CRC press.

Brière, D., & Traverse, P. (1993). AIRBUS A320/A330/A340 electrical flight controls - a family of fault-tolerant systems. In *Twenty-third international symposium on fault-tolerant computing* (pp. 616–623). IEEE.

Bundesstelle für Flugunfalluntersuchung. (2004). *Investigation Report 5x011-0/02 Serious Incident 3 December 2002 near Munich Airbus Industry / A300-600* https://www.bfu-web.de/EN/Publications/FinalReports/2002/Report_02_5X011-0_Munic_A300.pdf

Bundesstelle für Flugunfalluntersuchung. (n.d.). *Investigation Reports*. https://www.bfu-web.de/EN/Publications/InvestigationReport/reports_node.html.

Burdun, I. Y. (1998). *The intelligent situational awareness and forecasting environment (The S.A.F.E. concept): A case study*. (SAE Technical Paper 981223). https://doi.org/10.4271/981223

Burdun, I. Y., & Parfentyev, O. M. (1999). Fuzzy situational tree-networks for intelligent flight support. *Engineering Applications of Artificial Intelligence, 12*(4), 523-541. https://doi.org/10.1016/S0952-1976(99)00012-3

Bureau d'Enquêtes et d'Analyses. (n.d.). *Investigation Reports*. https://www.bea.aero/no_cache/les-enquetes/evenements-notifies/

Bureau d'Enquêtes et d'Analyses. (1989). *Final report concerning the accident which occurred on Jun 26th 1988 at Mulhouse-Habsheim to the Airbus A320, registered F-GFKC*. https://reports.aviation-safety.net/1988/19880626-0_A320_F-GFKC.pdf

Bureau d'Enquêtes et d'Analyses. (1993). *Official Report of The Commission of Investigation Into the Accident on 20 January 1992 Near Mont Sainte*

Odile (Bas-Rhin) of the Airbus A.320 Registered F-GGED Operated by Air Inter. https://reports.aviation-safety.net/1992/19920120-0_A320_F-GGED.pdf

Bureau d’Enquêtes et d’Analyses. (2008). *Incident survenu le 21 novembre 2007 secteur sud de la France, croisière (FL) 410 à l’Airbus 330-202 immatriculé F-WWKK exploité par Airbus, livraison à Air Mauritius*. https://bea.aero/fileadmin/documents/docspa/2007/f-kk071121/pdf/f-kk071121_06.pdf

Bureau d’Enquêtes et d’Analyses. (2010). *Accident on 27 November 2008 off the coast of Canet-Plage to the Airbus A320-232 registered D-AXLA operated by XL Airways Germany*. https://bea.aero/fileadmin/documents/docspa/2008/d-la081127.en/pdf/d-la081127.en.pdf

Bureau d’Enquêtes et d’Analyses. (2016). *Serious incident on 24 May 2011 during descent to Kuala Lumpur Airport (Malaysia) to the Dassault Falcon 7X registered HB-JFN operated by Jet Link AG*. https://bea.aero/fileadmin/uploads/tx_elydbrapports/hb-n110525.en_01.pdf

Bureau of Aircraft Accidents Archives. (n.d.). *Bureau of Aircraft Accidents Archives (B3A)*. https://www.baaa-acro.com

Bureau of Aircraft Accidents Archives. (2019). *Crash of a Douglas DC-9-15 off Margarita Island: 11 killed*. https://baaa-acro.com/crash/crash-douglas-dc-9-15-margarita-island-11-killed

Burns, C. P. (2000). *Analysing accident reports using structured and formal methods* (Unique ID: 2000-3554). [Doctoral dissertation, University of Glasgow]. http://theses.gla.ac.uk/id/eprint/3554

Canadian Transportation Accident Investigation and Safety Board Act. (1989) (S.C. 1989, c. 3).

Carriger, J. F., & Barron, M. G. (2011). Minimizing risks from spilled oil to ecosystem services using influence diagrams: The Deepwater Horizon spill response. *Environmental Science & Technology*, *45*(18), 7631–7639. https://doi.org/10.1021/es201037u

CAST ICAO Common Taxonomy Team. (2013). *Aviation occurrence categories: Definitions and usage notes (version 4.6)*. http://www.intlaviationstandards.org/Documents/OccurrenceCategoryDefinitions.pdf

Charmaz, K. (2017). The Power of Constructivist Grounded Theory for Critical Inquiry. *Qualitative Inquiry*, *23*(1), 34–45. https://doi.org/10.1177/1077800416657105

Charmaz, K. (2017a). Constructivist grounded theory. *The Journal of Positive Psychology*, *12*(3), 299–300. https://doi.org/10.1080/17439760.2016.1262612

Chou, C.-D. (1991). *Cockpit task management errors: A design issue for intelligent pilot-vehicle interfaces.* [Doctoral dissertation, Oregon State University]. https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/qn59q7812

Civil Aviation (Investigation of Air Accidents and Incidents) Regulations 2018 (UK). (2018). https://www.legislation.gov.uk/uksi/2018/321/contents/made

Clark, H. H. (1996). *Using language*. Cambridge University Press. https://doi.org/10.1017/CBO9780511620539

Comisión de Investigación de Accidentes e Incidentes de Aviación Civil. (2003). *Incident of the aircraft Airbus A-340-313, registered EC-GPB, at Salamanca Airport on 8 November 2002* (Technical Report No. IN-076/2002). Ministerio de Fomento. https://www.mitma.gob.es/recursos_mfom/2002_076_in_eng1_1.pdf

Committee on Transportation and Infrastructure. (1997). *Status of the "Investigation of the Crash of TWA 800" and the Proposal concerning the*

"Death On The High Seas Act". (Transcript 42-320CC). http://www.house.gov/transportation

Cook, R. I. (1998). *How complex systems fail*. Cognitive Technologies Laboratory, University of Chicago*.* https://www.adaptivecapacitylabs.com/HowComplexSystemsFail.pdf

Cook, R. I. (2012). *Lectures on the study of cognitive work*. The Royal Institute of Technology, Huddinge, Sweden.

Cook, R. I., & Long, B. A. (2021). Building and revising adaptive capacity sharing for technical incident response: A case of resilience engineering. *Applied Ergonomics*, *90*, 103240. https://doi.org/10.1016/j.apergo.2020.103240

Cooper, H. M., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis* (2nd ed). Russell Sage Foundation.

Cox Jr, L. A. (2015). *Breakthroughs in decision science and risk analysis*. John Wiley & Sons.

Curry, Renwick E. (1985, May 1). *The Introduction of New Cockpit Technology: A Human Factors Study* (NASA Technical Memorandum TM-86659). https://ntrs.nasa.gov/citations/19850019217

Damos, D. (1991). *Multiple task performance*. CRC Press. https://doi.org/10.1201/9781003069447

Degani, A., & Wiener, E. L. (1997). Procedures in complex systems: The airline cockpit. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *27*(3), 302–312. https://doi.org/10.1109/3468.568739

Dekker, S. (2002). Reconstructing human contributions to accidents: The new view on error and performance. *Journal of Safety Research*, *33*(3), 371–385.

Dekker, S. W. A. (2003). Illusions of explanation: A critical essay on error classification. *The International Journal of Aviation Psychology, 13*(2), 95-106. https://doi.org/10.1207/S15327108IJAP1302_01

Dekker, S. W. A. (2004). *Ten questions about human error: A new view of human factors and system safety*. CRC Press.

Dekker, S. W. A. (2005). We need new accident models. In D. Harris & H. Muir (Eds.), *Contemporary issues in human factors and aviation safety*, (pp. 181-198). Aldershot: Ashgate Publishing Co.

Dekker, S., & Hollnagel, E. (Eds.). (1999). *Coping with computers in the cockpit* (1st ed.). Routledge. https://doi.org/10.4324/9780429460609

Dekker, S. W. A., & Woods, D. D. (2002). MABA-MABA or Abracadabra? Progress on Human-Automation Co-ordination. *Cognition, Technology & Work*, *4*(4), 240–244. https://doi.org/10.1007/s101110200022

Dekker, S., & Pitzer, C. (2016). Examining the asymptote in safety progress: A literature review. *International Journal of Occupational Safety and Ergonomics*, *22*(1), 57–65. https://doi.org/10.1080/10803548.2015.1112104

Denzin, N. K., & Lincoln, Y. S. (Eds.). (2011). *The Sage handbook of qualitative research* (4th ed.). Sage Publications.

Diez, F. J., & Druzdzel, M. J. (2006). Reasoning under Uncertainty. In L. Nadel (Ed.), *Encyclopedia of Cognitive Science*. https://doi.org/10.1002/0470018860.s00028

Direction general de l'armement (DGA). (1994). *Commission d'enquete sur l'accident survenu le 30 Juin 1994 a Toulouse-Blagnac (31) a l'Airbus A330 No. 42 d'Airbus Industrie immatricule FWWKH Rapport Preliminaire*. www.rvs.uni-bielefeld.de/publications/Incidents/DOCS/ComAndRep/A330-Toulouse/Rapport.html

Directorate of Aviation Safety Bundeswehr. (2021). *Aircraft accident investigation report summary Global 5000 Berlin-Schönefeld, 16 April 2019*. German Military Aviation Authority. https://www.isasi.org/Documents/library/technical-papers/2021/

Dismukes, K., Berman, B. A., & Loukopoulos, L. D. (2007). *The limits of expertise: Rethinking pilot error and the causes of airline accidents*. Ashgate Publishing Ltd.

Edwards, W. (1954). The theory of decision making. *Psychological Bulletin*, *51*(4), 380-417. https://doi.org/10.1037/h0053870

Edwards, W., Miles, R. F., & Von Winterfeldt, D. (Eds.). (2007). *Advances in decision analysis: From foundations to applications*. Cambridge University Press. https://doi.org/10.1017/CBO9780511611308.002

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532–550.

Eisenhardt, K. M. (1991). Better Stories and Better Constructs: The Case for Rigor and Comparative Logic. *Academy of Management Review*, *16*(3), 620–627. https://doi.org/10.5465/amr.1991.4279496

Eisenhardt, K. M. (2021). What is the Eisenhardt Method, really? *Strategic Organization*, *19*(1), 147–160. https://doi.org/10.1177/1476127020982866

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32–64. https://doi.org/10.1518/001872095779049543

Endsley, M. R., & Kiris E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors, 37*(2), 381-394.

Eriksson, A., & Stanton, N. A. (2015). When communication breaks down or what was that? The importance of communication for successful coordination in complex systems. In *6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015*. UK: Elsevier B.V. https://doi.org/10.1016/j.promfg.2015.07.501

European Aviation Safety Agency. (2008). *Notice of proposed amendment (NPA) 2008-20 Draft decision of the executive director of the European Aviation Safety Agency – Flight testing*. (NPA No 2008-20). https://www.easa.europa.eu/sites/default/files/dfu/NPA%202008-20.pdf

European Aviation Safety Agency. (2012). *Notice of proposed amendment (NPA) 2012-08 Draft opinion of the European Aviation Safety Agency – Maintenance check flights (MCFs)*. (NPA No 2012-08). https://www.easa.europa.eu/sites/default/files/dfu/NPA%202012-08.pdf

European Aviation Safety Agency. (2017). *Opinion No 01/2017 Maintenance check flights*. (Opinion 01/2017). https://www.easa.europa.eu/en/document-library/opinions/opinion-012017

European Aviation Safety Agency. (2018). *Flight test operations manual guide* (Initial Issue, April 2018). https://www.easa.europa.eu/sites/default/files/dfu/FTOM%20Guide.pdf

Evans, J. St. B. T., Over, D. E., & Manktelow, K. I. (1993). Reasoning, decision making and rationality. *Cognition*, *49*(1), 165–187. https://doi.org/10.1016/0010-0277(93)90039-X

Favre, C. (1994). Fly-by-wire for commercial aircraft: The Airbus experience. *International Journal of Control*, *59*(1), 139–157. https://doi.org/10.1080/00207179408923072

Federal Aviation Act of 1958, Pub. L. No. 85-726, 72 Stat. (1958).

Federal Aviation Administration. (n.d.). *Lessons learned from civil aviation accidents*. (Boeing Model 737-3B7 USAir Flight 427). http://lessonslearned.faa.gov/index.cfm

Federal Aviation Administration. (2002). *Guidance addressing operational check flights following maintenance of air carrier aircraft*. (FAA Order 8300.10, Appendix 4, FSAW 02-12).

Federal Aviation Administration. (2008a). *Non-routine flight operations (NRFO)*. (InFO 08032). https://www.faa.gov/othervisit/aviationindustry/airlineoperators/airlinesafety/non-routine-flight-operations-nrfo

Federal Aviation Administration. (2008b). *Review of flight data recorder data from non-revenue flights*. (SAFO 08024). https://www.faa.gov/sites/faa.gov/files/other_visit/aviation_industry/airline_operators/airline_safety/SAFO08024.pdf

Federal Aviation Administration. (2016). *Non-Revenue Flight Procedures*. (InFO16006). https://www.faa.gov/sites/faa.gov/files/other_visit/aviation_industry/airline_operators/airline_safety/InFO16006.pdf

Federal Aviation Regulation Part 25 – Airworthiness Standards Transport Category Airplanes, 14 C.F.R. § 25 (2023). https://www.ecfr.gov/current/title-14/chapter-I/subchapter-C/part-25

Federal Aviation Regulation Part 91 – General Operating and Flight Rules, 14 C.F.R. § 91 (2023). https://www.ecfr.gov/current/title-14/chapter-I/subchapter-F/part-91

Ferroff, C. V. (2014). *Culture and its impact on flight deck management*. [Doctoral dissertation, Griffith University]. https://doi.org/10.25904/1912/2016

Fischhoff, B. (1975). Hindsight is not equal to foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, *1*(3), 288-299. https://doi.org/10.1037/0096-1523.1.3.288

Fischhoff, B. (1982). Debiasing. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 422-444). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511809477.032

Fischhoff, B., & Beyth, R. (1975). I knew it would happen: Remembered probabilities of once future things. *Organizational Behavior and Human Performance*, *13*(1), 1–16. https://doi.org/10.1016/0030-5073(75)90002-1

Flight Safety Foundation. (n.d.). *Aviation safety net database*. https://aviation-safety.net/database/databases.php

Flight Safety Foundation. (2011a). *Functional check flight compendium*. https://flightsafety.org/toolkits-resources/functional-check-flights/fcf_compendium/

Flight Safety Foundation. (2011b). *Functional check flight symposium*. https://flightsafety.org/toolkits-resources/functional-check-flights/

Flight Simulation Systems. (n.d.). *Flight Simulation Systems LLC Library*. https://www.fss.aero/accident-reports/browse_type.php?type=flight_type.

Flyvbjerg, B. (2011). Case study. *The Sage handbook of qualitative research* (4th ed.) (pp. 301–316). Thousand Oaks, CA: Sage.

Funk, K. (1991). Cockpit Task Management: Preliminary Definitions, Normative Theory, Error Taxonomy, and Design Recommendations. *The International Journal of Aviation Psychology*, *1*(4), 271–285. https://doi.org/10.1207/s15327108ijap0104_2

Funk, K., Lyall, B., Wilson, J., Vint, R., Niemczyk, M., Suroteguh, C., & Owen, G. (1999). Flight Deck Automation issues. *The International Journal of Aviation Psychology*, *9*(2), 109–123. https://doi.org/10.1207/s15327108ijap0902_2

Gabinete de Prevenção e Investigação de Acidentes com Aeronaves e de Acidentes Ferroviários. (2020). *Aileron cables reversal during maintenance actions and consequent loss of control in flight.* (Report No. AC_08/ACCID/2018_RF). http://www.gpiaa.gov.pt

Gephart, R. P., Jr. (1993). The textual approach: Risk and blame in disaster sensemaking. *Academy of Management Journal*, *36*(6), 1465-1514. ProQuest One Academic.

Gephart, R. P., & Saylors, R. (2020). Qualitative Designs and Methodologies for Business, Management, and Organizational Research. In R. P. Gephart & R. Saylors, *Oxford Research Encyclopedia of Business and Management*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190224851.013.230

Glaser, B., & Strauss, A. (1967). Grounded theory: The discovery of grounded theory. *Sociology the journal of the British sociological association*, *12*(1), 27-49.

Goodrick, D. (2019). Comparative Case Studies. In P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug, & R. A. Williams (Eds.), *SAGE Research Methods*

Foundations. SAGE Publications Ltd. https://doi.org/10.4135/9781526421036849021

Harris, D. (2003). The human factors of fully automatic flight, *Measurement and Control*, *36*(6),184-187. https://doi.org/10.1177/002029400303600605

Hawkins, F. H., & Orlady, H. W. (1993). *Human factors in flight* (Revised Second). Burlington, VT USA: Ashgate Publishing.

Heider, F. (1958). The naive analysis of action. In F. Heider, *The psychology of interpersonal relations* (pp. 79–124). John Wiley & Sons Inc. https://doi.org/10.1037/10628-004

Helmreich, R. L., Klinect, J. R., & Wilhelm, J. A. (1999). Models of threat, error, and CRM in flight operations. In *Proceedings of the tenth international symposium on aviation psychology* (pp. 677–682).

Herrera, I.A., & Woltjer, R. (2010). Comparing a multi-linear (STEP) and systemic (FRAM) method for accident analysis. *Reliab. Eng. Syst. Saf., 95*(12), 1269-1275. https://doi.org/10.1016/j.ress.2010.06.003

Hitt, E. F., & Mulcare, D. (2000). Fault-tolerant avionics. In C. R. Spitzer (Ed.), *Digital avionics handbook*. CRC press.

Hoffman, R. R., & Lintern, G. (2006). Eliciting and representing the knowledge of experts. In K. A. Ericsson, N. Charness, P. J. Feltovich, & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 203–222). Cambridge University Press. https://doi.org/10.1017/CBO9780511816796.012

Hollnagel, E. (2002). Understanding accidents-from root causes to performance variability. In *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants* (pp. 1-6). https://doi.org/10.1109/HFPP.2002.1042821

Hollnagel, E. (2011). *Resilience engineering in practice: A guidebook*. Farnham, Surrey: Ashgate Publishing Limited.

Hollnagel, E. (2013). The Fukushima disaster - systemic failures as the lack of resilience. *Nuclear Engineering and Technology, 45*(1), 13-20. https://doi.org/10.5516/NET.03.2011.078

Hollnagel, E. (2016). *Barriers and accident prevention*. Routledge. https://doi.org/10.4324/9781315261737

Hollnagel, E, & Goteman, O. (2004). The functional resonance accident model. *Proceedings of Cognitive System Engineering in Process Plant*, *2004*, 155–161.

Hollnagel, E, & Woods, D. D. (1983). Cognitive systems engineering: new wine in new bottles. *International Journal of Man-Machine Studies*, *18*(6), 583–600.

Hollnagel, E, & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press.

Hollnagel, E, Woods, D. D., & Leveson, N. (2006). *Resilience engineering: Concepts and precepts*. Ashgate Publishing, Ltd.

Holloway, C. M., & Johnson, C. W. (2004). *Distribution of causes in selected US aviation accident reports between 1996 and 2003*. https://ntrs.nasa.gov/api/citations/20040085800/downloads/20040085800.pdf

Holloway, C. M., & Johnson, C. W. (2006). Why system safety professionals should read accident reports. In *1st IET International Conference on System Safety* (Vol. 2006, pp. 325–331). https://doi.org/10.1049/cp:20060233

Hopkins, A. (2019). *Organising for safety*. Wolters Kluwer.

Howard, R. A. (1988). Decision Analysis: Practice and Promise. *Management Science*, *34*(6), 679–695. https://doi.org/10.1287/mnsc.34.6.679

Howard, R. A., & Matheson, J. E. (2005). Influence Diagram Retrospective. *Decision Analysis*, *2*(3), 144–147. https://doi.org/10.1287/deca.1050.0050

Inagaki, T. (2005). Design of Human Interactions with Smart Machines: Lessons Learned from Aircraft Accidents. *The 4th IARP/IEEE RAS/EURON, Keynote Lecture*.

Inagaki, T. (2010). Traffic systems as joint cognitive systems: Issues to be solved for realizing human-technology coagency. *Cognition, Technology & Work*, *12*(2), 153–162. https://doi.org/10.1007/s10111-010-0143-6

Inagaki, T. (2014). Human coagency for collaborative control. *Cybernics: Fusion of Human, Machine and Information Systems*, 235–265.

Independent Safety Board Act of 1974, 49 U.S.C.A. § 11 (2006). https://www.ntsb.gov/legal/Pages/ntsb_statute.aspx

International Air Transport Association. (2016). *Environmental Factors Affecting Loss of Control In-Flight: Best Practice for Threat Recognition & Management*. https://www.iata.org/en/programs/safety/operational-safety/loss-of-control-inflight/

International Air Transport Association. (2020). *Safety Report 2019* (Edition 56). https://www.iata.org/en/publications/safety-report/references/

International Civil Aviation Organization. (1993). *Human Factors Digest No. 7 Investigation of Human Factors in Accidents and Incidents*. (Circular 240-AN/144).

International Civil Aviation Organization. (2011). *Manual of aircraft accident and incident investigation Part III Investigation*. (DOC 9756).

International Civil Aviation Organization. (2013). *Manual of Evidence-based Training.* (DOC 9995).

International Civil Aviation Organization. (2014). *Manual on aeroplane upset prevention and recovery*. (DOC 10011*).*

International Civil Aviation Organization. (2015). *Manual of aircraft accident and incident investigation Part I Organization and planning*. (DOC 9756*).*

International Civil Aviation Organization. (2019). *Safety Report - Universal Safety Oversight Audit Programme - Continuous monitoring approach results 1 January 2016 to 31 December 2018*. https://www.icao.int/safety/cmaforum/Pages/default.aspx

International Civil Aviation Organization. (2018). *Safety Management Manual (SMM)*. (DOC 9859).

International Civil Aviation Organization. (2020a). *Annex 13 to the International Convention on Civil Aviation—Aircraft Accident and Incident Investigation*.

International Civil Aviation Organization. (2020b). *Manual of aircraft accident and incident investigation Part IV Reporting*. (DOC 9756*).*

International Civil Aviation Organization. (2022). *ICAO Safety Report* (No. 2022 ed.). https://www.icao.int/safety/Pages/Safety-Report.aspx

International Federation of Air Line Pilots' Associations. (2019). *Non-routine operations.* (Position Paper No. 19POS01). https://www.ifalpa.org/publications/library/non-routine-operations-2926

Javaux, D. (1998). Explaining Sarter & Woods' Classical Results. The Cognitive Complexity of Pilot-Autopilot Interaction on the Boeing 737-EFIS. In *Proceedings of Human Error, Safety, and System Development (HESSD'98)* (pp. 62–77).

Johnson, C. W. (1997a). Proving Properties of Accidents. In C. M. Holloway & K. J. Hayhurst (Eds.), *Fourth NASA Langley Formal Methods Workshop*.

Johnson, C. W. (1997b). Reasoning about Human Error and System Failure for Accident Analysis. In S. Howard, J. Hammond, & G. Lindgaard (Eds.), *Human-Computer Interaction INTERACT '97: IFIP TC13 International Conference on Human-Computer Interaction, 14th18th July 1997, Sydney, Australia* (pp. 331–338). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-35175-9_54

Johnson, C. W. (2003). *A Handbook of Incident and Accident Reporting*. Glasgow University Press.

Johnson, C. W., & Holloway, C. M. (2003). A survey of logic formalisms to support mishap analysis. *Reliability Engineering & System Safety*, *80*(3), 271–291. https://doi.org/10.1016/S0951-8320(03)00053-X

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kanouse, D. E., & Hanson, L. R., Jr. (1987). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 47–62). Lawrence Erlbaum Associates, Inc.

Keeney, R. L. (1982). Decision analysis: An overview. *Operations Research*, *30*(5), 803–838. https://doi.org/10.1287/opre.30.5.803

Klein, G. A. (2017). *Sources of power: How people make decisions* (20th Anniversary Edition). Cambridge, MA: MIT Press.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. In *Proceedings of the human factors society annual meeting* (Vol. 30, pp. 576–580). Sage Publications.

Klein, G. A., & Klinger, D. (1991). Naturalistic Decision Making. *Human Systems IAC Gateway*, *11*(3).

Klein, G. A., Orasano, J., Calderwood, R., & Zsambok, C. E. (1993). *Decision Making in Action: Models and Methods*. Ablex Publishers.

Klein, G. A., Woods, D. D., Bradshaw, J. M., Hoffman, R. R., & Feltovich, P. J. (2004). Ten Challenges for Making Automation a "Team Player" in Joint Human-Agent Activity. *IEEE Intelligent Systems*, *19*(06), 91–95. https://doi.org/10.1109/MIS.2004.74

Klein, G., Feltovich, P. J., Bradshaw, J. M., & Woods, D. D. (2005). Common Ground and Coordination in Joint Activity. In W. B. Rouse & K. R. Boff (Eds.), *Organizational Simulation* (pp. 139–184). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/0471739448.ch6

Klein, G., Moon, B., & Hoffman, R. R. (2006a). Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intelligent Systems*, *21*(4), 70–73. https://doi.org/10.1109/MIS.2006.75

Klein, G., Moon, B., & Hoffman, R. R. (2006b). Making Sense of Sensemaking 2: A Macrocognitive Model. *IEEE Intelligent Systems*, *21*(5), 88–92. https://doi.org/10.1109/MIS.2006.100

Klein, G., & Wright, C. (2016). Macrocognition: From theory to toolbox. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00054

Kosko, B. (1986). Fuzzy cognitive maps. *International Journal of Man-Machine Studies*, *24*(1), 65–75. https://doi.org/10.1016/S0020-7373(86)80040-2

Kosko, B. (1988). Hidden patterns in combined and adaptive knowledge networks. *International Journal of Approximate Reasoning*, *2*(4), 377–393. https://doi.org/10.1016/0888-613X(88)90111-9

La Porte, T. R. (1996). High reliability organizations: Unlikely, demanding and at risk. *Journal of Contingencies and Crisis Management*, *4*(2), 60–71.

Ladkin, P. (2000). Causal reasoning about aircraft accidents. *Computer Safety, Reliability and Security*, 344–360. https://doi.org/10.1007/3-540-40891-6_30

Ladkin, P., & Loer, K. (1999). Formalism helps in describing accidents. *Gateway to the New Millennium 18th Digital Avionics Systems Conference Proceedings.* (Cat. No.99CH37033).

Landa, A. H., Szabo, I., Brun, L. L., Owen, I., & Fletcher, G. (2011). An Evidence-Based Approach to Scoping Reviews. *The Electronic Journal of Information Systems Evaluation*, *14*(1).

Lande, K. (2016). Aircraft controllability and primary flight displays - Every link is important. In *ISASI 2016*. Reykjavik, Iceland: International Society of Air Safety Investigators.

Landman, A., Groen, E. L., van Paassen, M. M. (René)., Bronkhorst, A. W., & Mulder, M. (2017). Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *59*(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Le Coze, J. C. (2022). The ‘new view’ of human error. Origins, ambiguities, successes and critiques. *Safety Science*, *154*, 105853. https://doi.org/10.1016/j.ssci.2022.105853

Leveson, N. (1995). *Safeware*. Addison-Wesley.

Leveson, N. (2001). *Evaluating accident models using recent aerospace accidents, Part 1: Event-based models*. Massachusetts Institute of Technology, Engineering Systems Division. http://hdl.handle.net/1721.1/102726

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science, 42*(4), 237–270. https://doi.org/10.1016/S0925-7535(03)00047-X

Leveson, N. (2011). Applying systems thinking to analyze and learn from events. *Safety Science, 49*(1), 55–64. https://doi.org/10.1016/j.ssci.2009.12.021

Leveson, N. (2012). *Engineering a safer world: Systems thinking applied to safety*. MIT Press.

Leveson, N. (2014). Applying systems thinking to aviation psychology. In M. A. Vidulich & P. S. Tsang & J. Flach (Eds.), *Advances in aviation psychology* (Vol. 1) (pp. 17-27). Ashgate.

Leveson, N. (2016). Rasmussen's legacy: A paradigm change in engineering for safety. *Applied Ergonomics, 59*(B), 581-591. https://doi.org/10.1016/j.apergo.2016.01.015

Leveson, N. (2020). Are you sure your software will not kill anyone? *Communications of the ACM*, *63*(2), 25–28. https://doi.org/10.1145/3376127

Leveson, N., Dulac, N., Marais, K., & Carroll, J. (2009). Moving beyond normal accidents and high reliability organizations: A systems approach to safety in complex systems. *Organization Studies*, *30*(2-3), 227–249. https://doi.org/10.1177/0170840608101478

Levy, Y., & J. Ellis, T. (2006). A Systems Approach to Conduct an Effective Literature Review in Support of Information Systems Research. *Informing Science: The International Journal of an Emerging Transdiscipline*, *9*, 181–212. https://doi.org/10.28945/479

Lewis, D. (2008). Causation. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 632–638). Cambridge University Press. https://doi.org/10.1007/s10503-014-9315-5

Libyan Civil Aviation Authority. (2020). *Embraer ERJ 170–REG - 5A-SOC 2013-12-06*. https://caa.gov.ly/eng/wp-content/uploads/2016/pdf/5A-SOC-2013-Final.pdf

Lunenburg, F., & Irby, B. (2008). *Writing a Successful Thesis or Dissertation: Tips and Strategies for Students in the Social and Behavioral Sciences*. Corwin Press. https://doi.org/10.4135/9781483329659

Maitlis, S., & Sonenshein, S. (2010). Sensemaking in Crisis and Change: Inspiration and Insights From Weick (1988). *Journal of Management Studies*, *47*(3), 551–580. https://doi.org/10.1111/j.1467-6486.2010.00908.x

Malmquist, S., & Leveson, N. (2019). Investigating accidents in highly automated systems - Systemic problems identified through analysis of Air France 447. In *ISASI 2019*. International Society of Air Safety Investigators.

Martin, W. L. (2014). *Pathological behaviours in pilots during unexpected critical events: The effects of startle, freeze and denial on situation outcome.* [Doctoral dissertation, Griffith University]. https://doi.org/10.25904/1912/225

Maxwell, J. A. (2009). Designing a Qualitative Study. In L. Bickman (Ed.), *The SAGE Handbook of Applied Social Research Methods*. SAGE.

Maxwell, J. A. (2013). *Qualitative Research Design* (3rd ed.). SAGE Publications Ltd.

McKimmie, T., & Szurmak, J. (2002). Beyond Grey Literature: How Grey Questions Can Drive Research. *Journal of Agricultural & Food Information*, *4*(2), 71–79. https://doi.org/10.1300/J108v04n02_06

McNamee, P., & Celona, J. (2007). *Decision analysis for the professional*. SmartOrg, Incorporated.

McRuer, D., & Graham, D. (2004). Flight control century: Triumphs of the systems approach. *Journal of Guidance Control and Dynamics*, *27*(2), 161–173. https://doi.org/10.2514/1.4586

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Moshansky, V. P. (1992). *Commission of Inquiry into the Air Ontario crash at Dryden, Ontario*. https://publications.gc.ca/pub?id=9.699856&sl=0

Mosier, K. L. (2008). Technology and “naturalistic” decision making: Myths and realities. In *Naturalistic decision making and macrocognition* (pp. 41–54). Ashgate Aldershot, UK.

Mosier, K. L., Skitka, L. J., Heers, S., & Burdick, M. (1998). Automation bias: Decision making and performance in high-tech cockpits. *The International Journal of Aviation Psychology*, *8*(1), 47–63. https://doi.org/10.1207/s15327108ijap0801_3

Mumford, M. D., McIntosh, T., & Mulhearn, T. (2018). Using Cases to Understand Expert Performance: Method and Methodological Triangulation. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (2nd ed., pp. 291–309). Cambridge University Press. https://doi.org/10.1017/9781316480748.017

Nakamura, D., McKenney, D., & Railsback, P. (2013). *Operational use of flight path management systems*. Federal Aviation Administration. https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/afx/afs/afs400/parc/parc_reco

National Transportation Safety Board. (n.d.). *Aviation accident database and synopses*. https://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx

National Transportation Safety Board. (1973). *Aircraft accident report Eastern Air Lines, L-1011, N310EA, Miami Florida, December 29, 1972.* (No. AAR-73-14). https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR7314

National Transportation Safety Board. (1994). *A review of flightcrew-involved major accidents of US air carriers, 1978 through 1990*. (No. SS9401). https://www.ntsb.gov/safety/safety-studies/Pages/SS9401.aspx

National Transportation Safety Board. (1996). *Aircraft Accident Report Controlled Flight Into Terrain American Airlines Flight 965 Boeing 757-223, N651AA Near Cali, Colombia December 20, 1995* (Final Report DCA96RA020). Aeronautica Civil of The Republic of Colombia.

National Transportation Safety Board. (1997). *Aircraft accident report Uncontrolled flight into terrain ABX Air (Airborne Express) Douglas DC-8-63, N827AX Narrows, Virginia December 22, 1996* (No. AAR-97/05). https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR9705.pdf

National Transportation Safety Board. (2001). *Safety Recommendation A-96-094*. https://www.ntsb.gov/safety/safety-recs/_layouts/ntsb.recsearch/Recommendation.aspx?Rec=A-96-094

National Transportation Safety Board. (2002). *Accident record Turkish Airlines Flight 5904, Boeing 737-400, Ceyhan, Adana, Turkey* (No. DCA99RA053).

National Transportation Safety Board. (2003). *Safety Recommendation A-97-050*. https://data.ntsb.gov/carol-main-public/sr-details/A-97-050

National Transportation Safety Board. (2005). *Aviation accident Final report Fort Lauderdale, FL 03/11/2004, 0653 EST Airbus Industrie A300F4-605R* (No. MIA04IA056).

National Transportation Safety Board. (2007). *Crash of Pinnacle Airlines Flight 3701 Bombardier CL-600-2B19, N8396A Jefferson City, Missouri October 14, 2004*. (No. DCA05MA003). https://www.ntsb.gov/investigations/Pages/DCA05MA003.aspx

National Transportation Safety Board. (2012). *Crash During Experimental Test Flight Gulfstream Aerospace Corporation GVI (G650), N652GD Roswell, New Mexico April 2, 2011*. (No. DCA11MA076). https://www.ntsb.gov/investigations/Pages/DCA11MA076.aspx

National Transportation Safety Board. (2015). *Final report Dallas, TX 11/20/2014, 2000 CST Boeing 737 7H4 Ground collision.* (No. DCA15CA025A).

Nelson, H. (2008a). Be prepared: Check flights and aircraft upsets. In *61st annual International Air Safety Summit*. Honolulu, Hawaii: Flight Safety Foundation.

Nelson, P. S. (2008b). *A STAMP analysis of the Lex Comair 5191 accident*. [Unpublished master's thesis]. Lund University, Sweden.

Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, *2*(2), 175–220.

Norman, D. A. (1986). Cognitive engineering. In *User Centered System Design* (pp. 31-61). Lawrence Erlbaum Association.

Norman, D. A. (1990). The 'problem' with automation: Inappropriate feedback and interaction, not "over-automation." *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, *327*(1241), 585–593.

Nouvel, D., Travadel, S., & Hollnagel, E. (2007). Introduction of the concept of functional resonance in the analysis of a near-accident in aviation. In *33rd ESReDA Seminar: Future challenges of accident investigation*. Ispra, Italy.

Nuclear Regulatory Commission. (2000). *Technical Basis and Implementation Guidelines for A Technique for Human Event Analysis (ATHEANA)* (NUREG-1624 Rev. 1). https://www.nrc.gov/reading-rm/doc-collections/nuregs/staff/sr1624/index.html

Oliver, N., Calvard, T., & Potočnik, K. (2017). Cognition, Technology, and Organizational Limits: Lessons from the Air France 447 Disaster. *Organization Science*, *28*(4), 729–743. https://doi.org/10.1287/orsc.2017.1138

Orasanu, J., & Connolly, T. (1993). The reinvention of decision making. In G. A. Klein, J. Orasanu, R. Calderwood, & C. E. Zsambok (Eds.), *Decision Making in Action: Models and Methods*, (pp. 3–20). Ablex Publishing.

Orasanu, J., & Fischer, U. (1997). Finding decisions in natural environments: The view from the cockpit. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 343–357). Lawrence Erlbaum Associates, Inc.

Orasanu, J., & Martin, L. (1998). Errors in Aviation Decision Making: A factor in accidents and incidents. *HESSD '98*, 100–107.

Orasanu, J., Martin, L., & Davison, J. (1998). *Errors in Aviation Decision Making: Bad decisions or bad luck?* (Report 20020063485). National Aeronautics and Space Administration. https://ntrs.nasa.gov/api/citations/20020063485

Parasuraman, R. E., & Mouloua, M. E. (Eds.). (1996). *Automation and human performance: Theory and applications*. Lawrence Erlbaum Associates, Inc.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on*

Systems, Man, and Cybernetics - Part A: Systems and Humans, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Patton, M. Q. (2002). *Qualitative research & evaluation methods*. (3rd ed.). Sage Publications.

Pearl, J. (2000). *Causality, models, reasoning, and inference*. Cambridge University Press.

Perrow, C. (1981). Normal accident at Three Mile Island. *Society*, *18*(5), 17–26. https://doi.org/10.1007/BF02701322

Perrow, C. (1999). *Normal accidents: Living with high risk technologies - updated edition.* Princeton University Press.

Perrow, C. (2004). A personal note on normal accidents. *Organization & Environment*, *17*(1), 9–14. https://doi.org/10.1177/1086026603262028

Perrow, C. (2011). *The Next Catastrophe: Reducing Our Vulnerabilities to Natural, Industrial, and Terrorist Disasters*. Princeton University Press.

Pillemer, D. B. (1998). What is remembered about early childhood events? *Clinical Psychology Review*, *18*(8), 895-913.

Pinet, J., & Bück, J-C. (Eds.) (2013). *Dossier 37 - Dealing with unforeseen situations in flight*. Air and Space Academy.

Poprawa, S. (2015). Maintenance test flying an accident waiting to happen? *The Aeronautical Journal*, *119*(1216), 781–790. https://doi.org/10.1017/S0001924000010824

Raiffa, H. (1968). *Decision analysis: Introductory lectures on choices under uncertainty.* Addison Wesley.

Rankin, A., Woltjer, R., & Field, J. (2016). Sensemaking following surprise in the cockpit a re-framing problem. *Cognition, Technology & Work*, *18*(4), 623–642. https://doi.org/10.1007/s10111-016-0390-2

Rasmussen, J. (1987). *Information processing and human-machine interaction. An approach to cognitive engineering*. North-Holland.

Rasmussen, J. (1990). Human error and the problem of causality in analysis of accidents. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 449–462. https://doi.org/10.1098/rstb.1990.0088.

Rasmussen, J. (1997). Risk management in a dynamic society: A modelling problem, *Safety Science 27*(2), 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0

Rasmussen, J., & Svedung I. (2000). *Proactive risk management in a dynamic society.* Swedish Rescue Services Agency.

Reason, J. (1990). *Human error*. Cambridge University Press.

Reason, J. (2008). *The human contribution*. Ashgate.

Regulation (EC) No 216/2008. (2008). Regulation (EC) No 216/2008 of the European Parliament and of the Council of 20 February 2008 on common rules in the field of civil aviation and establishing a European Aviation Safety Agency. https://www.easa.europa.eu/en/document-library/regulations/regulation-ec-no-2162008

Regulation (EU) No 996/2010. (2010). Regulation (EU) No 996/2010 of the European Parliament and of the Council of 20 October 2010 on the investigation and prevention of accidents and incidents in civil aviation. https://www.easa.europa.eu/en/document-library/regulations/regulation-eu-no-9962010

Regulation (EU) 2019/1384. (2019). Commission Implementing Regulation (EU) 2019/1384 of 24 July 2019 amending Regulations (EU) No 965/2012 and (EU) No 1321/2014 as regards the use of aircraft listed on an air operator certificate for non-commercial operations and specialised operations, the establishment of operational requirements for the conduct of maintenance check flights, the establishment of rules on non-commercial operations with reduced cabin crew on board and introducing editorial updates concerning air operations requirements. http://data.europa.eu/eli/reg_impl/2019/1384/oj

Roberts, K. H. (1990). Managing high reliability organizations, *California Management Review 32*(4), 101–113. https://doi.org/10.2307/41166631

Roland, H. E. & Moriarty, B. (1990). *System Safety Engineering and Management*. Wiley.

Rogovin, M. (1980). *Three Mile Island: A report to the commissioners and to the public*. Nuclear Regulatory Commission, Special Inquiry Group. United States. https://doi.org/10.2172/5395798

Roth, W.-M. (2018). Autopsy of an airplane crash: A transactional approach to forensic cognitive science. *Cognition, Technology & Work*, *20*(2), 267–287. https://doi.org/10.1007/s10111-018-0465-3

Sagan, S. D. (1993). *The limits of Safety - Organizations, accidents, and nuclear weapons*. Princeton University Press.

Sagan, S. D. (2004). Learning from normal accidents. *Organization & Environment*, *17*(1), 15–19. https://doi.org/10.1177/1086026603262029

Sarter, N. B. (1994). *" Strong, silent, and 'out-of-the-loop'": Properties of advanced (cockpit) automation and their impact on human-automation interaction* (Order No. 9517075). [Doctoral dissertation, The Ohio State University]. ProQuest Dissertations and Theses Global.

Sarter, N. B., & Woods, D. D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system. *The*

International Journal of Aviation Psychology, *2*(4), 303–321. https://doi.org/10.1207/s15327108ijap0204_5

Sarter, N. B., & Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. *Human Factors*, *39*(4), 553–569.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation surprises. In *Handbook of Human Factors & Ergonomics* (pp. 1926-1943). John Wiley and Sons.

Shappell, S., Hackworth, C., Boquet, A., & Wiegmann, D. A. (2007). Human Error and Commercial Aviation Accidents: An Analysis Using the Human Factors Analysis and Classification System. *Human Factors*, 16.

Shattuck, L. G., & Miller, N. L. (2006). Extending naturalistic decision making to complex organizations: A dynamic model of situated cognition. *Organization Studies*, *27*(7), 989–1009. https://doi.org/10.1177/0170840606065706

Sheridan, T. B. (1988). Chapter 8 - task allocation and supervisory control. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 159–173). Amsterdam: North-Holland. https://doi.org/10.1016/B978-0-444-70536-5.50013-0

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. MIT Press.

Shrivastava, P., Mitroff, I. I., Miller, D., & Miclani, A. (1988). Understanding Industrial Crises. *Journal of Management Studies*, *25*(4), 285-303. https://doi.org/10.1111/j.1467-6486.1988.tb00038.x

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics 69*, 99–118. https://doi.org/10.2307/1884852

Simulation of Upset Recovery in Aviation. (2013, January 18). *Final report summary - SUPRA (Simulation of Upset Recovery in Aviation)*. (Record number 56069). https://cordis.europa.eu/project/id/233543/reporting

Sklet, S. (2002). *Methods for accident investigation*. Norwegian University of Science and Technology. https://dvikan.no/ntnu-studentserver/reports/accident.pdf

Sklet, S. (2004). Comparison of some selected methods for accident investigation. *Journal of Hazardous Materials*, *111*(1-3), 29–37. https://doi.org/10.1016/j.jhazmat.2004.02.005

Sloman, S. A. (2008). When Explanations Compete: The Role of Explanatory Coherence on Judgments of Likelihood. In J. E. Adler & L. J. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 343–352). Cambridge University Press.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology*, *28*(1), 1–39. https://doi.org/10.1146/annurev.ps.28.020177.000245

Snook, S. A. (2000). *Friendly fire: The accidental shootdown of U.S. Black Hawks over Northern Iraq*. Princeton, N.J: Princeton University Press.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Strauch, B. (2002). Normal accidents: Yesterday and today. In C. Johnson (Ed.), *Proceedings of the first workshop on the investigation and reporting of incidents and accidents (IRIA 2002)* (pp. 10-18). Department of Computing Science, University of Glasgow, Scotland.

Strauch, B. (2017). *Investigating human error: Incidents, accidents and complex systems* (2nd edition). Boca Raton: CRC Press.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.

Strauss, A., & Corbin, J. (1990). Qualitative research. *Grounded Theory; SAGE Publications Ltd.: New York, NY, USA*.

Thomas, J., & Malmquist, S. (2016). Learning from Accidents that are a Consequence of Complex Systems. In *ISASI 2016*. Reykjavik, Iceland: International Society of Air Safety Investigators.

Transport Accident Investigation Commission. (n.d.). *Aviation Occurrence Reports*. https://www.taic.org.nz/inquiries.

Transport Safety Investigation Act 2003. (2003). (Cth).

Underwood, P., & Waterson, P. (2013). Systemic accident analysis: Examining the gap between research and practice. *Accident Analysis & Prevention*, *55*, 154–164. https://doi.org/10.1016/j.aap.2013.02.041

Van der Schaaf, T. W. (1995). Human recovery of errors in man-machine systems. *IFAC Proceedings Volumes*, *28*(15), 71–76.

Vaughan, D. (1990). Autonomy, Interdependence, and Social Control: NASA and the Space Shuttle Challenger. *Administrative Science Quarterly*, *35*(2), 225–257.

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

Veillette, P. R. (2009). Danger lurks in non-routine flights: Post-maintenance test flights contain an extraordinary number of additional risks. *Business and Commercial Aviation*, *105*(5), 36-42.

Von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd rev. ed.). Princeton University Press.

Weick, K. E. (1979). *The social psychology of organizing*. McGraw-Hill.

Weick, K. E. (1987). Organizational culture as a source of high reliability, *California Management Review 29*(2), 112-127. https://doi.org/10.2307/41165243

Weick, K. E. (1988). Enacted Sensemaking in Crisis Situations. *Journal of Management Studies*, *25*(4), 305–317. https://doi.org/10.1111/j.1467-6486.1988.tb00039.x

Weick, K. E. (1990). The vulnerable system: An analysis of the Tenerife air disaster. *Journal of Management*, *16*(3), 571–593. https://doi.org/10.1177/014920639001600304

Weick, K. E. (1993). The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. *Administrative Science Quarterly*, *38*(4), 628–652. https://doi.org/10.2307/2393339

Weick, K.E. (1998). Foresights of failure: An appreciation of Barry Turner. *Journal of Contingencies and Crisis Management, 6*, 72-75. https://doi.org/10.1111/1468-5973.00072

Weick, K. E. (2004). Normal accident theory as frame, link, and provocation. *Organization & Environment*, *17*(1), 27–31. https://doi.org/10.1177/1086026603262031

Weick, K. E. (2010). Reflections on Enacted Sensemaking in the Bhopal Disaster. *Journal of Management Studies*, *47*(3), 537–550. https://doi.org/10.1111/j.1467-6486.2010.00900.x

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the Process of Sensemaking. *Organization Science*, *16*(4), 409–421. https://doi.org/10.1287/orsc.1050.0133

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance* (4th ed.). Psychology Press.

Wiegmann, D. A., & Shappell, S. A. (2000). *The Human factors analysis and classification system (HFACS)* (No. DOT/FAA/AM-00/7). Federal Aviation Administration. https://www.tc.faa.gov/its/worldpac/techrpt/am00-7.pdf

Wiegmann, D. A., & Shappell, S. A. (2001). Human error perspectives in aviation. *The International Journal of Aviation Psychology*, *11*(4), 341–357. https://doi.org/10.1207/S15327108IJAP1104_2

Wienen, H. C. A., Bukhsh, F. A., Vriezekolk, E., & Wieringa, R. J. (2017). *Accident Analysis Methods and Models—A Systematic Literature Review* (CTIT Technical Report TR-CTIT-17-04). Centre for Telematics and Information Technology (CTIT).

Wiener, E. L. (1989). *Human factors of advanced technology (glass cockpit) transport aircraft*. (NASA Contractor Report 177528). Human Systems Integration Division. National Aeronautics and Space Administration. https://human-factors.arc.nasa.gov/publications/HF_AdvTech_Aircraft.pdf

Wiener, E. L., Chute, R. D., & Moses, J. H. (1999). *Transition to glass: Pilot training for high-technology transport aircraft*. (ADA532148). https://apps.dtic.mil/sti/citations/ADA532148

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, *23*(10), 995–1011. https://doi.org/10.1080/00140138008924809

Wiener, E. L., & Nagel, D. C. (1988). *Human factors in aviation*. Academic Press.

Wiggins, M. W., & Glass, R. (2013). Improving the Accessibility of Investigation Reports Using Safety Factor Maps. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*(1), 20–24. https://doi.org/10.1177/1064804612463694

Woltjer, R., & Hollnagel, E. (2007). The Alaska Airlines flight 261 accident: A systemic analysis of functional resonance. *2007 International Symposium on Aviation Psychology*, 763-768. https://corescholar.libraries.wright.edu/isap_2007/4

Woods, David (2003, October 29). *Creating foresight: How resilience engineering can transform NASA's approach to risky decision making.* Committee on Commerce, Science and Transportation.

Woods, D. D. (1996). Decomposing automation: Apparent simplicity, real complexity. In R. Parasuraman & M. Mouloua (Eds.), *Automation and Human Performance: Theory and Applications* (pp. 3–17). Lawrence Erlbaum.

Woods, D. D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability Engineering & System Safety*, *141*, 5–9. https://doi.org/10.1016/j.ress.2015.03.018

Woods, D. D. (2018). The theory of graceful extensibility: Basic rules that govern adaptive systems. *Environment Systems and Decisions*, *38*(4), 433–457. https://doi.org/10.1007/s10669-018-9708-3

Woods, D. D. (2021). *Patterns in How People Think and Work*. Eurocontrol.

Woods, D. D., & Branlat, M. (2017). Basic Patterns in How Adaptive Systems Fail. In E. Hollnagel, J. Pariès, D. Woods, & J. Wreathall (Eds.), *Resilience Engineering in Practice* (1st ed., pp. 127–143). CRC Press. https://doi.org/10.1201/9781317065265-10

Woods, D. D., & Cook, R. I. (2017). Incidents: Markers of resilience or brittleness? In E. Hollnagel & D. D. Woods (Eds.), *Resilience Engineering* (pp. 69–76). CRC Press. https://doi.org/10.1201/9781315605685-10

Woods, D. D., & Hollnagel, E. (1987). Mapping cognitive demands in complex problem- solving worlds. *International Journal of Man-Machine Studies*, *26*, 257–275.

Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. CRC Press.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers, and hindsight*. Dayton University Research Institute.

Woods, D. D., Licu, T., Leonhardt, J., Rayo, M., Balkin, E. A., & Coponea, R. (2021). *Patterns in how people think and work*. Eurocontrol. https://skybrary.aero/sites/default/files/bookshelf/5987.pdf

Woods, D. D., & Sarter, N. B. (2000). Learning from Automation Surprises and “Going Sour” Accidents. In N. B. Sarter & R. Amalberti (Eds.), *Cognitive Engineering in the Aviation Domain* (pp. 327–353). Lawrence Erlbaum Associates Publishers

Xiao, Y., & Watson, M. (2019). Guidance on Conducting a Systematic Literature Review. *Journal of Planning Education and Research*, *39*(1), 93–112. https://doi.org/10.1177/0739456X17723971

Yeh, Y. C. (1996). Triple-triple redundant 777 primary flight computer. In *Aerospace Applications Conference, 1996. Proceedings., 1996 IEEE* (Vol. 1, pp. 293–307). IEEE. https://doi.org/10.1109/AERO.1996.495891

Yeh, Y. C. (1998). Design considerations in Boeing 777 fly-by-wire computers. In *High-Assurance Systems Engineering Symposium, 1998. Proceedings. Third IEEE International* (pp. 64–72). IEEE. https://doi.org/10.1109/HASE.1998.731596

Yeh, Ying Chin (2014, May 12). *Ultra-Reliable Fly-By-Wire Computers for Commercial Airplanes' Flight Controls Systems*. IEEE ComSoc Technical Committee on Communication Quality & Reliability. https://cqr2014.ieee-cqr.org/ETR-RT/Yeh_IEEE-ETR-RT-2014_Ultra%20Reliable_12May2014.pdf

Yin R. K. (2014). *Case study research: design and methods* (5th ed.). SAGE.

APPENDIX A – SEARCH CRITERIA

Table 10

NTSB Search

Database:	Aviation Accident Database and Synopses Query
Web address:	https://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx
Search criteria:	{EventDate between 01/01/1988 and 04/02/2020; Aircraft Category = AIRPLANE; and Amateur ≠ YES}
Additional filter:	{Engine Type = Geared Turbofan OR Turbo Jet OR Turbo Fan OR Blank; No of ENG > 1; SCHD ≠ SCHD"; Injuries ≤ 10; Purpose of FLT= Ferry OR Flight Test OR Instructional OR Positioning OR Unknown; Make AND Models=(SELECT only Western-built and MTOW>60,000 lb)}

Table 11

ATSB Search

Database:	ATSB National Aviation Occurrence Database
Web address:	https://www.atsb.gov.au/avdata
Search criteria:	FILTER by Aircraft Type = AEROPLANE; and Operation Type = Air Transport High Capacity – Check & Training OR Test & Ferry OR Other
Additional filter:	N/A

Table 12

TAIC Search

Database:	Aviation Safety Database
Web address:	https://www.taic.org.nz/inquiries
Search criteria:	If the aircraft was identified as a “Fixed-wing aircraft” AND “MTOW>27,000kg” AND “Commercial Jet”, then the report was evaluated for {Type of operation=Non-revenue Flight?}
Additional filter:	N/A

Table 13

BEA Search

Database:	Investigation Reports Menu
Web address:	https://www.bea.aero/no_cache/les-enquetes/evenements-notifies/
Search criteria:	FILTER by Aircraft Category = “Fixed Wing – Large Aeroplane” AND Class Flight = “Commercial Air Transport - Non-revenue Operations”
Additional filter:	N/A

Table 14

BFU Search

Database:	Interim Reports and Final Reports menu
Web address:	https://www.bfu-web.de/EN/Publications/Investigation%20Report/reports_node.html
Search criteria:	The “Interim Reports” AND “Final Reports” menu points under “Publications” were interrogated by SELECT Type of Aircraft=“>5700kg”
Additional filter:	N/A

Table 15

AAIB Search

Database:	Air Accident Investigation Branch reports
Web address:	https://www.gov.uk/aaib-reports
Search criteria:	FILTER by Aircraft Category =“Commercial - Fixed Wing” AND Report Type = “Formal Report” OR “Bulletin - Field Investigation” OR “Bulletin - Pre-1997” OR Bulletin - Correspondence Investigation” OR “Foreign Report” OR “Special bulletin”
Additional filter:	The following free text search results were performed: “post-maintenance” OR “ferry” OR “instructional” OR “test”

Table 16

AVH Search

Database:	Aviation Herald website articles
Web address:	http://avherald.com
Search criteria:	Simple text search field to yield articles with: “non-revenue flight” OR “post-maintenance” OR “check flight” OR “ferry flight” OR “positioning flight” in the article
Additional filter:	N/A

Table 17

ASN Search

Database:	Aviation Safety Network Database (Flight Safety Foundation)
Web address:	https://aviation-safety.net/wikibase/wikisearch.php
Search criteria:	Aircraft Category = Airplane AND Nature = “Test” OR “Training” OR “Ferry/positioning”
Additional filter:	N/A

Table 18

FSS Search

Database:	Flight Simulation Systems LLC database
Web address:	https://www.fss.aero/accident-reports/browse_type.php?type=flight_type
Search criteria:	The data was searched by Type of Flight = Engineering Test Flight OR Ferry OR Training
Additional filter:	N/A

APPENDIX B - NRFO CATALOGUE

Table 19

Non-routine Accidents and Incidents (1988-2021)

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
AAIB	Incident	13 Jul 2021	London Luton	United Kingdom	None	None	A319-111	Operational Check
BFU	Incident	7 Jul 2021	Cologne-Bonn	Germany	None	Minor	B737-400	Positioning
AAIB	Incident	26 May 2021	Lasham	United Kingdom	None	None	A319-111	Ferry
NBAAI	Incident	21 May 2021	Hurghada	Egypt	None	None	B737-8H6	Positioning
TSBC	Incident	22 Dec 2020	Tucson, AZ	United States	None	None	B737-8 MAX	Ferry
AAIB	Incident	16 Jun 2020	Doncaster	United Kingdom	None	None	A321-231	Ferry
AAIB	Incident	26 Feb 2020	London Gatwick	United Kingdom	None	Minor	A321-211	Positioning

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
BEA	Accident	10 Jan 2020	Antalya	Turkey	None	Destroyed	A321-231	Positioning
AIB	Incident	6 Feb 2019	Lahore	Pakistan	Incident	None	A321-272N	Ferry
TSBC	Incident	12 Dec 2018	Paris Orly	France	Incident	None	B737-800	Ferry
GPIAAF	Accident	11 Nov 2018	Alverca	Portugal	Minor	Substantial	E190	Operational Check
AAIB	Incident	10 Oct 2018	Manchester	United Kingdom	None	None	B737-8K5	Ferry
TSBC	Incident	22 Jun 2018	St John’s	Canada	Incident	None	B737-200	Ferry
OJK	Accident	28 Feb 2018	Tallin	Estonia	Minor	Substantial	A320-214	Instructional
BEA	Incident	7 Feb 2018	Paris Orly	France	None	None	B737-800	Ferry
TSIB	Incident	16 Nov 2017	Singapore	Singapore	Incident	Minor	B737-700	Positioning

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
NTSB	Incident	7 Nov 2017	Seattle-Tacoma, WA	United States	None	Moderate	A330-243	Positioning
AAIU	Incident	9 Sep 2017	Sofia	Bulgaria	Incident	None	A320-231	Ferry
AAIB	Incident	22 Jul 2017	Newcastle	United Kingdom	Incident	Minor	B767-343	Positioning
SUST	Incident	7 Nov 2016	Geneva	Switzerland	Incident	Minor	A319-111	Ferry
TSBC	Incident	23 Aug 2016	Montreal	Canada	Incident	None	A319-114	Ferry
AAIB	Incident	13 Jul 2016	Edinburgh	United Kingdom	Incident	None	B767-322	Ferry
SUST	Incident	12 Apr 2015	Zurich	Switzerland	Incident	None	B737-500	Ferry
NTSB	Incident	10 Apr 2015	Victorville, CA	United States	Incident	Minor	B707-338C	Positioning
CIAA	Incident	29 Mar 2015	Tarapoto	Peru	Incident	None	B737-230	Ferry

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
NTSB	Accident	20 Nov 2014	Dallas, TX	United States	Non-Fatal	Minor	B737-7H4	Check Flight
ATSB	Incident	12 Mar 2014	Mildura	Australia	Incident	None	A320-232	Ferry
LYCAA	Accident	6 Dec 2013	Tripoli	Libya	Incident	Substantial	E170	Check Flight
AAIB	Incident	10 Apr 2013	Prestwick	United Kingdom	Incident	Minor	A320-214	Instructional
AAIB	Incident	1 Mar 2013	London Gatwick	United Kingdom	Incident	Minor	B737-377	Ferry
NTSB	Accident	24 Feb 2013	Tunica, MS	United States	Non-Fatal	None	E190	Positioning
TSBC	Incident	1 Nov 2012	Gander	Canada	Incident	None	B757-200	Ferry
AAIB	Incident	7 Aug 2012	North Sea	United Kingdom	Incident	None	B757-2K2	Check Flight
AIB	Accident	1 May 2012	Jeddah	Saudi Arabia	Incident	Substantial	A300B4-605R	Positioning

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
NTSB	Accident	6 May 2011	Greenville, MS	United States	Non-Fatal	Substantial	B737-800	Positioning
NTSB	Accident	30 Mar 2011	Dayton, OH	United States	Non-Fatal	Substantial	B737-301	Positioning
BEA	Incident	12 Jan 2011	Paris Orly	France	Incident	None	MD-83	Ferry
BEA	Incident	10 Jan 2011	Montpelier	France	Incident	None	B737-300	Ferry
SACAA	Incident	10 Jan 2011	Hoedspruit	South Africa	None	Substantial	B737-200	Positioning
SIAF	Incident	11 Dec 2010	Arkhangelsk	Russia	Incident	None	A330-302	Ferry
ATSB	Incident	16 May 2010	Adelaide	Australia	Incident	None	E190	Positioning
TSBC	Incident	28 Feb 2010	St John’s	Canada	Incident	None	B757-200	Ferry
AAIB	Incident	23 Feb 2010	Jersey Airport	United Kingdom	Incident	Substantial	E195	Ferry

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
BEA	Incident	20 Dec 2009	Reims FIR	France	Incident	None	MD-83	Positioning
AAIB	Incident	29 May 2009	London	United Kingdom	Incident	None	B737-73V	Check Flight
MOT	Accident	29 Apr 2009	Kwilu, Bandundu	Congo	Fatal(7)	Destroyed	B737-200	Ferry
AAIB	Incident	12 Jan 2009	Norwich	United Kingdom	Incident	None	B737-73V	Check Flight
BEA	Accident	27 Nov 2008	Perpignan	France	Fatal(7)	Destroyed	A320-232	Check Flight
EDGAC	Accident	30 Aug 2008	Latacunga	Ecuador	Fatal(3)	Destroyed	B737-200	Ferry
NTSB	Incident	4 Jul 2008	Jeddah	Saudi Arabia	Incident	Minor	B747-300	Ferry
NTSB	Accident	14 Dec 2007	New York, NY	United States	Non-Fatal	Substantial	E190	Positioning
BEA	Incident	21 Nov 2007	South France	France	Incident	None	A330-202	Check / Acceptance

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
AAIB	Incident	6 Sep 2007	Belfast	United Kingdom	Incident	None	BAE146-300	Positioning
ATSB	Incident	4 Feb 2007	Melbourne	Australia	Incident	None	B747-338	Positioning
AAIB	Incident	8 Nov 2006	Bristol	United Kingdom	Incident	Minor	B767-2Q8	Ferry
AAIB	Incident	22 Oct 2006	London Stansted	United Kingdom	Incident	None	B757-204	Check Flight
AAIB	Incident	1 Nov 2005	Not recorded	United Kingdom	Incident	None	B737-36N	Check Flight
NTSB	Accident	22 Oct 2005	San Antonio, TX	United States	Non-Fatal	None	B727-225	Positioning
AAIB	Incident	30 Jul 2005	Nottingham	United Kingdom	Incident	Minor	B757-236	Instructional
ATSB	Incident	6 Dec 2004	Rockhampton	Australia	Incident	None	F100	Positioning
NTSB	Incident	8 Oct 2004	Anchorage, AK	United States	Incident	Minor	MD-11F	Ferry

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
NTSB	Incident	11 Mar 2004	Fort Lauderdale, FL	United States	Incident	Minor	A300F4-605R	Operational Check
DAAIB	Incident	24 Jan 2004	Copenhagen	Denmark	Incident	Minor	B757-236	Positioning
BFU	Incident	3 Dec 2002	Munich	Germany	Minor	None	A300-600	Operational Check
CIAIAC	Accident	8 Nov 2002	Salamanca	Spain	Non-Fatal	Substantial	A340-313	Check Flight
ARAIC	Accident	26 Jun 2002	Shimoji-Shima	Japan	Minor	Moderate	B767-200	Instructional
NTSB	Incident	27 Jul 2001	OSHKOSH, WI	United States	Incident	None	B727-2Q6	Ferry
NTSB	Incident	25 Nov 2000	NEWARK, NJ	United States	Incident	None	MD-11	Check Flight
AAIU	Incident	12 May 2000	Dublin	Ireland	Incident	Minor	B747-212B	Check Flight
AIBT	Accident	3 Feb 2000	Mwanza	Tanzania	Non-Fatal	Substantial	B707-351C	Positioning

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
ASC	Incident	2 Sep 1999	Taipei	Taiwan	Incident	Substantial	B747SP	Instructional
ARIB	Accident	7 Apr 1999	Ceyhan	Turkey	Fatal(6)	Destroyed	737-4Q8	Positioning
AIBN	Accident	22 Feb 1998	Kaduna	Nigeria	Incident	Destroyed	B737-2K3	Instructional
AAIU	Incident	19 Dec 1997	Shannon	Ireland	Incident	None	MD-82	Check Flight
TAIC	Incident	16 Dec 1997	Sydney	Australia	Incident	Minor	B767-200	Positioning
DGCA	Incident	5 Dec 1997	Montevideo	Uruguay	Incident	None	B707-320	Positioning
NTSB	Accident	21 Nov 1997	SYRACUSE, NY	United States	Non-Fatal	Substantial	DC-9-15	Positioning
NTSB	Accident	22 Dec 1996	NARROWS, VA	United States	Fatal(6)	Destroyed	DC-8-63F	Check Flight
NTSB	Incident	14 Jun 1996	CHICAGO, IL	United States	Incident	None	B727-30C	Positioning

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
AAIBI	Accident	9 Jun 1996	Rasht Airport	Iran	Fatal(4)	Destroyed	B727-286	Instructional
NTSB	Accident	20 Mar 1996	WILMINGTON, OH	United States	Non-Fatal	Substantial	DC-8-62	Positioning
CIPAA	Accident	4 Feb 1996	Ascuncion	Paraguay	Fatal(22)	Destroyed	DC-8-55F	Instructional
NTSB	Incident	29 Oct 1995	SAN FRANCISCO, CA	United States	Incident	None	B737-500	Check Flight
AAIB	Incident	22 Oct 1995	Bournemouth	United Kingdom	Incident	None	B737-236	Check Flight
NTSB	Accident	16 Feb 1995	KANSAS CITY, MO	United States	Fatal(3)	Destroyed	DC-8-63	Ferry
NTSB	Accident	28 Jan 1995	BELLEVILLE, MI	United States	Non-Fatal	Substantial	B747-238B	Positioning
AAIB	Accident	21 Dec 1994	COVENTRY, UK	United Kingdom	Fatal(5)	Destroyed	B737-2D6C	Ferry
NTSB	Incident	2 Nov 1994	CHICAG0, IL	United States	Incident	Minor	B747-251B	Positioning

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
NTSB	Incident	13 Oct 1994	DFW AIRPORT, TX	United States	Incident	None	MD-11	Ferry
AAIB	Accident	23 Jul 1994	London Gatwick	United Kingdom	Minor	None	A300B4	Positioning
AAIB	Incident	29 Apr 1994	London Heathrow	United Kingdom	Incident	Minor	Concorde	Check Flight
DGCAI	Accident	8 Mar 1994	Delhi International	India	Fatal(9)	Destroyed	B737-200	Instructional
NTSB	Incident	6 Jun 1993	ANCHORAGE, AK	United States	Incident	Minor	B737-200	Ferry
JIAAC	Accident	2 Apr 1993	Margarita Island	Venezuela	Fatal(11)	Destroyed	DC-9-15	Check Flight
NTSB	Accident	11 Mar 1993	SAINT LOUIS, MO	United States	Non-Fatal	Substantial	DC-9-31	Instructional
NTSB	Accident	20 Sep 1990	MARANA, AZ	United States	Fatal(1)	Destroyed	B707-321B	Ferry
None	Accident	11 Sep 1990	Newfoundland	Canada	Fatal(16)	Destroyed	B727-247	Ferry

Bureau	Event	Event Date	Location	Country	Injury	Damage	Model	Purpose
NTSB	Accident	2 Jun 1990	Unalakleet, AK	United States	Non-Fatal	Destroyed	B737-2X6C	Positioning
NTSB	Accident	8 Dec 1989	NEWARK, NJ	United States	Incident	Substantial	B727-31	Positioning
NTSB	Incident	31 Aug 1989	CHICAGO, IL	United States	Incident	None	MD-80	Positioning
BEA	Accident	26 Jun 1988	Mulhouse-Habsheim	France	Fatal(3)	Destroyed	A320-100	Demonstration

APPENDIX C - CHECK FLIGHT CAUSAL LABELS

Table 20

Causal Factor Distribution (1988-2021)

Year	Flight Crew Error	System / Component	Maintenance	Airworthiness / Operations	MCF regulations	Training	Design / Certification	Undetermined	Environment	Causes	Factors
2021		Cause (1)						Factor (1)		1	1
2018	Factor (1)	Cause (1)	Cause (1) Factor (5)	Factor (3)	Factor (1)		Factor (1)			2	11
2014	Cause (1)									1	0
2013	Cause (2)									2	0
2012		Cause (1)	Factor (2)							1	2
2009	Cause (2)	Cause (1)	Cause (1)	Factor (1)			Factor (2)			4	3
2009	Factor (2)		Cause (1) Factor (2)	Cause (1) Factor (1)	Factor (3)	Factor (1)	Factor (1)			2	10
2008	Cause (2) Factor (5)	Cause (1)	Factor (1)	Factor (1)	Factor (1)		Cause (1)			4	8
2007		Cause (1)		Cause (1)						2	0
2006		Cause (1)	Factor (1)			Factor (1)	Factor (1)			1	3
2005	Cause (1)	Cause (1)	Cause (1)			Factor (1)	Factor (1)			3	2

2004	Cause (1)									1	0
2002	Cause (3/2) Factor (2/1)		Cause (1)	Cause (1) Factor (4/1)		Factor (1)	Cause (2/1) Factor (1)		Factor (1)	7	9
2000		Cause (1)	Cause (1) Factor (1)				Factor (1)			2	2
1997		Cause (1)	Cause (1)							2	0
1996	Cause (2)	Factor (1)		Cause (1)		Factor (1)				3	2
1995	Cause (2/2)						Cause (3/1)			5	0
1994		Cause (1)					Cause (1) Factor (1)			2	1
Subtotal:	Cause (16) Factor (10)	Cause (11) Factor (1)	Cause (7) Factor (12)	Cause (4) Factor (10)	Factor (5)	Factor (5)	Cause (7) Factor (9)	Factor (1)	Factor (1)	45	54
									AVERAGE	*2.50*	*3.00*
									MEDIAN	*2*	*2*
									STD DEV	*1.62*	*3.76*

Table 21

Causal Labels (1988-2021)

Flight Crew Error	System / Component	Maintenance	Airworthiness / Operations	MCF regulations	Training	Design / Certification	Undetermined	Environment
failure>detect	failure>system> component	AMO> maintenance> error	Airline>AMO> supervision	OEM> instructions> test schedule	crew>training	OEM>design> system	undetermined	environment> wildlife
failure>attention	system> inoperative> alerting	AMO> maintenance> quality	NAA>AMO> oversight	NAA>guidance> crew	NAA>regulations >operations> training	design>failure> alerting		
failure>CRM		AMO> maintenance> management	Airline>crew> guidance	NAA>guidance> airworthiness	crew>training>abnormal operations	NAA> certification> alerting		

Flight Crew Error	System / Component	Maintenance	Airworthiness / Operations	MCF regulations	Training	Design / Certification	Undetermined	Environment
failure> communicate		AMO> maintenance> control	Airline> operations> oversight	OEM>guidance> airworthiness	Airline>training >simulator	design>system> logic		
decision>error		maintenance> systemic	management> operations			design>software >error		
deviation>SOP		OEM> instructions> maintenance	NAA> airworthiness> MEL			NAA> certification> software		
failure>control			Airline> operations> instructions			OEM>design> system> contamination		
decision> improvise			Airline> management> operations			OEM>design> failure> unexpected		

Flight Crew Error	System / Component	Maintenance	Airworthiness / Operations	MCF regulations	Training	Design / Certification	Undetermined	Environment
action> inadvertent			crew>fatigue					
failure>action			OEM> instructions> MMEL					
crew>action> inadequate			NAA> airworthiness> MMEL					
crew>action> inappropriate			Airline>AMO> coordination					
crew>action> inadvertent> stall								

Intentionally left blank.