TU Delft

Delft University of Technology

Document Version
Final published version

Citation (APA)
Mohrmann, J. F. W. (2026). *Airmanship 2.0: From human error to human asset*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:a8e72c2b-d021-448c-aa81-102a6acb2038

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.



AIR MAN SHIP 2.0

FROM HUMAN ERROR TO HUMAN ASSET

FREDERIK MOHRMANN

To my wife Marjolijn.

AIRMANSHIP 2.0

FROM HUMAN ERROR TO HUMAN ASSET

Dissertation

For the purpose of obtaining the degree of doctor
at the Delft University of Technology
by the authority of the Rector Magnificus,
Prof.dr.ir. H. Bijl,
chair of the Board for Doctorates,
to be defended publicly on
Wednesday, 9 December 2026, 12:30

by

Johannes Frederik Willem Mohrmann

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof.dr.ir. M. Mulder	Delft University of Technology, promotor
Prof.dr. S.W.A. Dekker	Griffith University, Australia, promotor

Independent members:

Prof.dr.ir D.A. Abbink	Delft University of Technology
Prof.dr. M.D. Pavel	Delft University of Technology
Dr. A. Adriaensen	Delft University of Technology
Prof.dr.ir J.A.A.M. Stoop	Lund University, Sweden
Dr. J. Blundell	Cranfield University, United Kingdom
Dr.ir C. Borst	Delft University of Technology, *reserve member*

Keywords: Aviation Safety, Human Factors, Complexity, Neurocognition, Airmanship, Training, Simulation, Pilots, Joint Cognitive System

Copyright © 2026 by Johannes Frederik Willem Mohrmann
ISBN 978-94-6518-543-9 (Paperback)
ISBN 978-94-6518-544-6 (PDF)
ISBN 978-94-6518-545-3 (EPub)

An electronic version of this dissertation is available at http://repository.tudelft.nl

Cover page is (ironically) AI generated by the author through Microsoft Bing image generation. The term "airmanship" refers to all the men and woman who take to the skies.

Summary

The past two decades have shown a shift in aviation safety and accident causation. While the safety performance of the aviation transport system is unprecedented, accidents are increasingly characterized by difficulty of flight crews asserting manual control, maintaining situation awareness and demonstrating effective decision making. To better understand the nature of this shift in accident causation and how it can be mitigated, this dissertation explores the future aviator's proficiency to manage modern aircraft systems and flight operations that increasingly feature complexity and opacity. By retrospectively studying the intricacies of human-machine interaction in modern, fourth generation aircraft, a prospective model for the future aviator can be defined based on effective and ineffective behaviors in modern aircraft.

The dissertation is the deliberate compilation of aviation operational and safety research performed in the period between 2013 and 2026. Two important foundations are the European Union (EU) Seventh Framework research program Manual Operation of Fourth Generation Airlines (Man4Gen) which was completed between 2012 and 2015, and a study investigating pilot resilience through neurocognitive training and mindfulness in 2018, among other research programs in the same domain. While these studies must be regarded in due consideration of the timeframe they were performed, the results and outcomes are still very relevant in the present, if not more so with Human-Artificial Intelligence (AI) Teaming (HAT) and Reduced Crew Operation (RCO) concepts being explored in the flight deck. As such, the overarching research question of this dissertation explores a possible—perhaps even critical—evolution of the pilot model to improve aviation safety in an era of increased complexity and opacity in modern flight operations.

Following the introduction in Chapter 1, Chapters 2 and 3 investigate flight crew responses to simulated opaque and ambiguous situations in modern aircraft. Both an exploratory and validation simulator study were performed in the respective chapters, introducing new behavioral analysis techniques that provide high-resolution data on crew behaviors and their safety performance. Findings from Chapter 2 indicated that when facing complexity and opacity, flight crews featured clear differences in behaviors (despite equal qualifications) that impacted flight safety. Safety-increasing behaviors include 1) a level of wariness and openness to (the likelihood of) opacity, 2) explicit, conscious valuation of the useability of procedures and automation and 3) making time for sensemaking and diagnostics to create an atmosphere conducive to navigating the opacity they are presented with. In contrast, safety-reducing behaviors included 1) contingency management (i.e., flight

diversion) without sufficient sensemaking, 2) lack of diagnostic structure and 3) both overreliance on or total avoidance of procedures and automation.

The validation study in Chapter 3 attempted to induce the desired behaviors with a novel problem-solving strategy and Risk Information System (RIS) designed around desired behaviors from the exploratory study, evaluating them with a new population and simulated scenario. Findings showed similar relations between behaviors and flight safety as in Chapter 2. Evaluating strategy and RIS treatments indicated limited strategy training effectiveness, while the RIS did feature some desired effects.

This leads to a transition from the retrospective, explorative domain to the prospective domain in the dissertation, beginning with the question how such adaptive behaviors could be embedded into the existing pilot model. The term pilot model is understood as the collection of roles and responsibilities attributed to the pilot in the flight deck socio-technical system. Chapter 4 explores the problematic dichotomy of reproducibility versus adaptability paradigms that strain the existing pilot model. A new, compound-nested pilot model is defined—Airmanship 2.0— which provides a framework in which both existing compliant and adaptive behaviors 1) are explicitly valued, 2) are protected against proficiency decay and 3) allow management of both transparent and opaque situations. This model is contrasted with hallmark commercial aviation accidents from the past two decades. Given the extensive exposure to transparent operations in the majority of operational flying, providing such a compound structure to the pilot model (and competency framework) may actively maintain the requisite proficiency of adaptive responses for infrequent but critical situations.

The Airmanship 2.0 restaging of sensemaking as the default response of flight crews invites a deeper understanding of how to train and operationalize it. Chapter 5 provides a further breakdown of the neurocognition of sensemaking with an extensive literature review and identifies three distinct key sensemaking archetypes: analytical-, heuristic-, and holistic sensemaking. Conceptual, analytical sensemaking is most appropriate for complex, unambiguous situations. A heuristic approach functions best for non-complex, yet ambiguous situations. Holistic sensemaking functions best in the context of complexity with ambiguity. Training for these three archetypes will differ. Conceptual sensemaking requires a down-regulation of the brain's Default Mode Network (DMN) possibly through focused attention mindfulness practices. Heuristic and holistic sensemaking require pre-developed expertise for intuitive cognitive functions to add value to the sensemaking process. Holistic sensemaking requires an upregulation of internal signals, possibly through open-awareness mindfulness practices.

The potential for directed neurocognitive training is explored in Chapter 5 with an experimental training and simulator research study (2018) investigating the impact of mindfulness training on crew response to ambiguous and opaque situations in modern aircraft. Trained crews featured higher scores for communication quality, emotional intelligence and calmness, and lower scores for focus. Outcome performance of trained and untrained crews exhibited some improvements in "startle & surprise management", "problem solving & decision making" and "workload management", which were hypothesized to be most positively impacted by the training. The mindfulness study experienced some resistance and decreased participation of pilots, possibly related to the personal nature of this mindset change, as well as confronting trainees with their tolerance for ambiguity.

In conclusion, this dissertation's investigation into flight crew cognitive control of complex systems and operations that feature opacity provides tentative evidence that existing aircrew standards and training may need to be revised to effectively mitigate new opacity-related flight safety threats. This may be achieved by operationally nesting compliance standards within a framework of sensemaking, individually tailored neurocognitive training and carefully targeted innovations in the cockpit Joint Cognitive System (JCS) to support sensemaking.

Contents

Summary **6**

Contents **9**

1 Introduction **1**

1.1 Introduction 2

1.2 The increasing cognitive distance 3

1.3 Research questions & approach 5

1.4 Research scope 7

1.5 Reading guide 8

2 Exploring behavior in opaque operations **3**

2.1 Introduction 4

2.2 Method 5

2.2.1 Theoretical Basis 5

2.2.2 Conceptual Framework 7

2.2.3 Experiment Design 10

2.3 Results 12

2.3.1 Performance Indicators 12

2.3.2 Correction Factors 15

2.3.3 Automation Awareness 16

2.3.4 Effect of Crew Combination 16

2.3.5 Combining Behavior with Performance 18

2.4 Discussion 19

3 Evaluating a new strategy for opaque operations **23**

3.1 Introduction 25

3.2 Man4gen project background 26

3.2.1 Summary of the Man4Gen strategy design 26

3.2.2 Extending the strategy with the Man4Gen Risk Information System.... 29
3.2.3 Man4Gen validation study 29
3.3 Method..... 30
3.3.1 Setup and scenario..... 30
3.3.2 Performance and behavior dependent measures..... 34
3.3.2.1 Crew performance measures 34
3.3.2.1.1 Crew behavioral measures..... 35
3.3.2.1.2 Behavior duration analysis method 36
3.3.2.1.3 Behavior sequence analysis method..... 40
3.3.2.2 Summary of measures..... 46
3.3.3 Study hypotheses 47
3.3.4 Data analysis methods..... 48
3.4 Results..... 50
3.4.1 Performance measures results and Analysis I..... 50
3.4.2 Behavioral measures results – Strategy phase level..... 52
3.4.3 Behavioral measures results – Principle level..... 56
3.4.4 Behavioral measures correlation – Analysis II..... 57
3.4.4.1 Between-duration measures correlation..... 58
3.4.4.2 Between-characterization measures correlation..... 59
3.4.4.3 Duration-characterization measures correlation..... 60
3.4.5 Training effectiveness results – Analysis III 62
3.4.6 Behavior versus performance correlation – Analysis IV..... 64
3.5 Reflection on hypotheses 66
3.5.1 Training effectiveness conclusions..... 66
3.5.2 Behavior-performance conclusions..... 67
3.6 Discussion 68
4 The Airmanship 2.0 concept 71
4.1 Introduction 72

4.2 The current flight deck control paradigm 73
4.2.1 Evolution of aircraft systems 73
4.2.2 Evolution of the pilot profession 74
4.2.3 Evolution of pilot-system interactions 75
4.2.4 The brittleness problem 76
4.3 Solving for the brittleness problem 79
4.3.1 The dichotomy of pilot control paradigms 79
4.3.2 The need for a singular pilot role 81
4.4 Airmanship 2.0: A compound pilot role 83
4.5 Seven anomalies of opacity 86
4.5.1 Air France 447 87
4.5.1.1 Synopsis (BEA, 2012) 87
4.5.1.2 Airmanship 2.0 reflection 87
4.5.2 Air Asia 8501 88
4.5.2.1 Synopsis (KNKT, 2015) 88
4.5.2.2 Airmanship 2.0 reflection 88
4.5.3 Asiana 214 89
4.5.3.1 Synopsis (NTSB, 2014) 89
4.5.3.2 Airmanship 2.0 reflection 90
4.5.4 West Air Sweden 206 90
4.5.4.1 Synopsis (SAIA, 2016) 90
4.5.4.2 Airmanship 2.0 reflection 90
4.5.5 Lion Air 610 & Ethiopian Airlines 302 91
4.5.5.1 Synopsis (KNKT, 2019; EAIB, 2022) 91
4.5.5.2 Airmanship 2.0 reflection 91
4.5.6 Quantas 32 92
4.5.6.1 Synopsis (ATSB, 2013) 92
4.5.6.2 Airmanship 2.0 reflection 92

4.6 Discussion ... 93

5 Making sense of sensemaking ... 97

5.1 Introduction ... 98

5.2 Safety in complexity ... 99

5.2.1 A bifurcated pilot role ... 99

5.2.2 The Airmanship 2.0 concept ... 100

5.3 The Neurocognition of sensemaking ... 102

5.3.1 Different neurocognitive pathways for sensemaking ... 102

5.3.1.1 Nuance 1: Multi-faceting of "intuition" ... 103

5.3.1.2 Nuance 2: Interactivity of sensemaking modalities ... 104

5.3.1.3 Nuance 3: The role of emotions and affect ... 105

5.3.1.4 Nuance 4: The nature of the situation ... 106

5.3.1.5 Nuance 5: Differences between experts and novices ... 106

5.3.1.6 Overview of sensemaking archetypes ... 107

5.3.2 Sensemaking neurocognition in the context of cockpit operations ... 109

5.4 Training sensemaking ... 112

5.4.1 Training novice pilots ... 112

5.4.2 Training expert pilots ... 114

5.4.3 Novice-expert sensemaking teaming ... 116

5.4.4 Possible evolutions of the training pipeline ... 117

5.5 Case study: Pilot mindfulness training ... 119

5.5.1 Study method ... 121

5.5.1.1 Mindfulness training and evaluation scenario ... 122

5.5.1.2 Mindfulness data ... 124

5.5.1.3 Flight safety performance data ... 125

5.5.2 Study results ... 126

5.5.2.1 Mindfulness survey results ... 127

5.5.2.2 Mindfulness biometric results ... 132

5.5.2.3 Flight Safety Performance Results 138
5.5.3 Study conclusions 141
5.6 Discussion 143
6 Discussion and reflections 147
6.1 Response to research questions 148
6.1.1 Identifying effective behaviors for opaque situations 149
6.1.2 Bridging adaptive behaviors with compliant behaviors 151
6.1.3 Training and operationalizing Airmanship 2.0 153
6.1.4 Overarching research question 155
6.2 Discussion 157
6.2.1 Evolving the flight deck socio-technical system 157
6.2.2 A shift in aviation safety monitoring 160
6.2.2.1 Discussion point 1 - Three new investigative lenses 160
6.2.2.2 Discussion point 2 - A shift in safety monitoring activities 162
6.2.3 Resilience of the aviation industry 163
6.3 Recommendations 168
6.4 Closing remarks 169
Acknowledgements 170
Appendix A 172
Appendix B 174
Appendix C 176
Appendix D 177
Appendix E 179
Appendix F 184
Appendix G 186
Appendix H 190
Appendix I 194
References 202

Abbreviations **226**

List of publications **230**

Curriculum vitae **231**

1
Introduction

This chapter introduces the challenges in aviation safety that have been the impetus for this thesis. It describes how the ideas put forward in this dissertation have evolved in the past decade but remain relevant for flight safety challenges of today and tomorrow. This chapter also provides an overview of the main research questions, the approach followed, assumptions made and a reading guide.

1.1 Introduction

In 2025 the worldwide commercial aviation system transported nearly 5 billion passengers with an accident rate of only 1.32 per million flights (IATA, 2026a). Compared to the year 2001, 2025 featured a 198% increase in number of sectors flown, while also achieving a 64% reduction in the number of accidents per million flights (IATA, 2002; IATA, 2026a). While the aviation industry's safety performance is approaching a 10^{-7} accident rate, it has been showing signs of stabilizing into a non-plus ultra-safe system for over a decade (Stoop, 2017; IATA 2025a; Airbus, 2025).

Looking at the year-over-year (YOY) improvement of the accident rate based on global safety data from Flight Safety Foundation & IATA (World Bank, 2026; FSF, 2026; IATA, 2002; IATA, 2006; IATA, 2011; IATA, 2014a; IATA, 2018; IATA, 2025b; IATA, 2026b; IATA, 2026c) shown in Figure 1.1, a few distinct periods can be identified. Since the mid 1990's the rate has featured consistent and positive annual improvements as well as a positive three-year moving average. However, around 2010 the YOY improvements reduced, showcasing a possible asymptotic limit of the safety performance.

This diminishing improvement has in part been attributed to specific flight deck dynamics in modern aircraft featuring *increasing cognitive distance* between modern crews' and aircraft systems and operations. These performance limiting dynamics are exacerbated by both the growing complexity of modern aircraft and operations, as well as a decrease in the requisite crew capacity to respond adaptively to such complexity (Woods, 2004; Dekker, 2009; Oster et al, 2013; IATA, 2013; Casner et al., 2014).

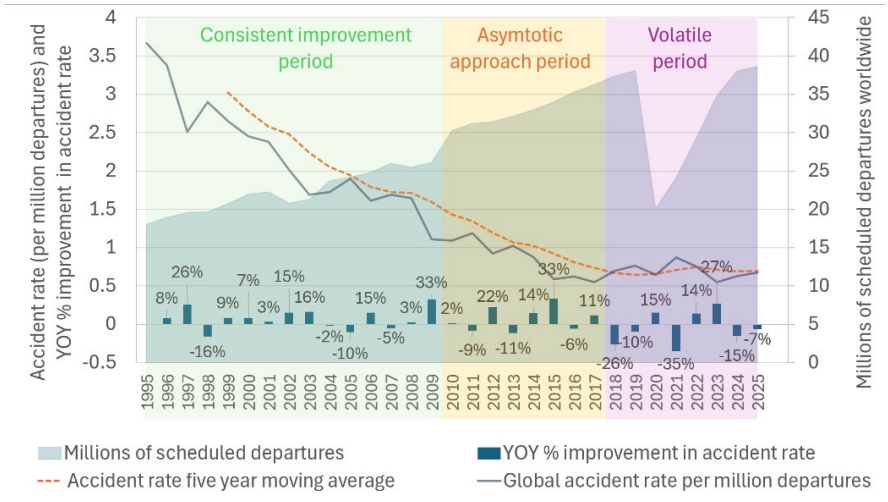


Figure 1.1. Year-over-year change in the accident rate per million sectors, five-year moving average and industry volume

This cognitive distance is not a wholly novel phenomenon: Sarter & Woods (1994), Mosier et al. (1998), Saurim & Carim (2012) and Manzey et al. (2012) described several performance-limiting mechanisms such as automation bias and automation complacency, with more recent studies highlighting similar phenomena in present aviation (Strauch, 2017; Kharoufah et al., 2018; Proctor & Van Zandt, 2018; Kelly & Efthymiou, 2019; Prinzel et al., 2024; Gago et al., 2025). Furthermore, this cognitive distance has the potential to grow even larger with increased system complexity and opacity that may result from AI-teaming (Eurocontrol, 2020; EASA, 2023; Lorrig & Daw, 2024; Kirwan 2025; Hörhager, 2025) as well as the advent of Reduced Crew Operations (RCO) (Malick & Gollnick, 2016; Sprengart et al., 2018; Schmidt & Stanton, 2020), or their combination (Minaskan et al., 2022). As a limiter of safety performance, this cognitive distance needs to be understood, acknowledged and mitigated.

In more recent years, the YOY improvement of the accident rate has continued in an unstable pattern [1], even showing a negative trend in the three-year moving average. This instability has mainly been attributed to significant industry disruptions including the COVID-19 pandemic, global conflicts and most recently an emerging oil crisis (2026), together with the industry's challenge to adapt and be resilient to such disruptions (IATA, 2025a). In terms of flight crew, similar underlying themes emerge including skill decay due to COVID-19 reduced operations (Curran and Pulver, 2024; IATA, 2026a, Li et al., 2024) and airline business volatility featuring greater pilot workforce turnover (EASA, 2019; Oliver Wyman, 2025). As such, the aviation industry is approaching a plateau of its safety performance, while flight crew proficiency continues to be a major subject of improvement, warranting a deeper investigation into the pilot role to identify safety barriers and improvement pathways.

1.2 The increasing cognitive distance

The original pilot in the early days of initial fixed wing flight were avid cyclists: their functional role was to contribute power to the aircraft system and manage basic lateral and vertical flight controls. Without the pilot, the aircraft would not take off and experience any form of controlled flight. Naturally the role has evolved over the years. No longer are pilots the source of power, but coordinators of multiple (types of) systems, managers of risk and compliance to prescribed standards and regulations.

[1] Note that a YOY graph will feature statistical rebound. In other words: a single unsafe year will feature a relative YOY decrease but also contribute to a *relative* increase for the following year. This is why a five-year moving average is valuable to observe the asymptotic development.

The introduction of "*safety by reproducibility"* through, for example, Standard Operating Procedures (SOP) and published procedures resulted in a marked increase in safety (Degani & Weiner, 1990). This was followed by an increase in flight deck automation which facilitated a shift from pilot-centric to automation-centric flight decks, where manual control made room for monitoring behaviors (Billings, 1996; Parasuraman & Riley, 1997). More recently, the industry's pursuit of increased operational reliability and predictability has resulted in fourth generation (jet) aircraft [2] feature complex system integrations, greater system autonomy beyond the scope of training programs (Casner et al., 2014; Haslbeck et al., 2014; Causse et al., 2025).

For over four decades, the pilot role has been increasingly seen through the lens of "human error" to identify and reduce aircrew variability, justifying a proliferation of automated systems as mitigations. Klein (2011) refers to this evolution as the Self-Reinforcing Complexity Loop. While each of these evolutions has improved the industry's safety performance, accident, incident and operational research indicate that this improvement may be at the expense of aircrew resilience and adaptability (Sarter & Woods, 1994, Mosier et al., 1998, Manzey et al., 2012; Casner et al, 2014; Strauch, 2017; Kelly & Efthymiou, 2019; Prinzel et al., 2024; Gago et al., 2025). Ironically, these may be precisely the (human) capacities required to manage the opacity present in more complex and autonomous systems. The heritage of the pilot role precludes that the pilot is always in full control, but evidence from incidents and accidents in the past decade reveals a cognitive distance (and possible control gap) in modern aircraft, and suggests that this (implicit) assumption may no longer hold.

As this cognitive distance may be the undesired side-effect of a shift from pilot-centric to automation-centric flight decks, more automation may not be the way to close it. Instead, the focus shifts to the human pilot's possible contribution to close the cognitive distance. While the pilot role has evolved to respond to automation in the past decades, including improved multi-crew coordination through Crew Resource Management (CRM) behaviors, compliance and automation management skills and most recently the shift to Competency Based Training & Assessment (CBTA) (ICAO, 2013; EASA, 2020), the specific dynamics between pilots and highly complex systems and the opacity they present crews, still features a knowledge gap. By means of a closer appraisal of the human-machine socio-technical system in modern, fourth generation aircraft, a commensurate next generation of aviator may need to be defined.

[2] Fourth generation commercial (jet) aircraft feature high levels of system integration and autonomy, digital flight controls and flight envelope protection. Refer to Airbus (2025) for further details about aircraft generations.

1.3 Research questions & approach

This dissertation explores the future aviator's proficiency to manage complex systems and their opacity through a combination of retrospection and prospection. By (retrospectively) studying the intricacies of human-machine interaction in modern, fourth generation aircraft, a (prospective) model for the future aviator can be defined based on effective and ineffective behaviors in such aircraft.

The dissertation is the deliberate compilation of aviation operational and safety research performed in the period between 2013 and 2026. Two important foundations are the EU Seventh Framework research program Manual Operation of Fourth Generation Airlines (Man4Gen) which was completed between 2012 and 2015, and a study investigating pilot resilience through neurocognitive training and mindfulness in 2018, among other research programs in the same domain. The Man4Gen program was a three-year program aiming to research and understand potential atrophy of aircrew manual and cognitive control in highly automated fourth generation aircraft. The research consortium was led by NLR, including DLR, Airbus, Boeing, Linköping University, University of Vienna, Medical University of Vienna, Global Training Aviation (GTA) and International Development of Technology (IDT) organizations. The author led, structured and organized two of three full-scale simulation studies within this program. The other mindfulness training research was conducted on behalf of a large commercial cargo operator, university and mindfulness research organization, where the author also led, designed and organized the simulation validation study.

While these studies must be regarded in due consideration of the timeframe they were performed, the results and outcomes are still relevant in the present, if not more so with Human-AI Teaming (HAT) and Reduced Crew Operation (RCO) concepts being explored in the flight deck. This leads to the overarching research question of this dissertation, focusing on a possible, perhaps even essential, evolution of the pilot role considering increasing automation and, critically, increased complexity and opacity related to it.

Overarching research question

How may an adapted pilot model improve the safety performance of highly automated commercial aircraft?

This research question is multi-faceted as it connects with safety research, human performance research as well as training and operationalization themes. Hence, a further breakdown into three research questions provides further structure to the dissertation.

A foundation for the dissertation lies in better understanding the cognitive distance between flight crews and the aircraft and operations, identifying those behaviors that are effective in managing complexity and opacity, and those behaviors that should be avoided. Both desired and undesired behaviors serve to provide contrast to the desired pilot role.

Research question 1

What specific aircrew behaviors contribute to or reduce control of highly automated commercial aircraft?

Chapters 2 and 3 address research question 1 by investigating flight crew responses to simulated opaque and ambiguous situations in modern aircraft. Both an exploratory and validation study were performed, introducing new behavioral analysis techniques that provide high-resolution data on crew behaviors and their safety performance, also addressing possible improvements in the broader flight deck socio-technical system, possibly in the form of a Joint Cognitive System (JCS). Both chapters pertain to studies performed in the Man4Gen EU research program (2012-2015).

With the set of desired and undesired behaviors known, the second research question is centered on how those behavioral insights can be formalized into a new pilot role. The core challenge in this question is less about formalization but rather lies in the effective harmonization of new behavioral standards with existing ones. In other words, advocating for the desired behaviors arising from research question 1 may not preclude that existing behavioral standards for prescribed and transparent operations are defunct. A careful appreciation of the total ensemble of requisite behavior is important.

Research question 2

How may an adapted pilot model integrate the desired behaviors with existing safety paradigms in commercial aviation?

Chapter 4 addresses research question 2 by transitioning from the retrospective, explorative domain to the prospective, anticipatory domain. By defining a new pilot model (*Airmanship 2.0*), the chapter investigates the harmonization of existing and novel safety paradigms intro a pilot model, contrasting this model with hallmark accidents which occurred in the past two decades.

With a conceptual pilot model in place, the third and final research question provides a prospective appreciation concerning the achievability of the concept, as this has bearing on which changes to the pilot model are feasible as well as the wider implications of those changes.

Research question 3

What are the training and operationalization implications of the newly proposed pilot model?

By a further breakdown of the neurocognitive processes related the new pilot model, Chapter 5 identifies both training pathways and operational implementation of the new pilot model. The identified training pathways are further explored through an experimental training and simulator research study (performed in 2018) investigating the impact of neurocognitive training (mindfulness) on crew response to ambiguous and opaque situations in modern aircraft.

1.4 Research scope

Research in human functioning in socio-technical systems is a broad field that can run deep in many dimensions, and it is not possible for this study to equally address all these dimensions. As such, this dissertation is scoped in three important ways.

The first scope is the operational context of modern, highly complex commercial transport aircraft. While human-machine interaction in less complex aircraft, military aircraft and smaller aircraft (including uncrewed arial vehicles) may show certain similarities in context and challenges, this study does not investigate these other platforms. Furthermore, the applicability of results from a multi-crew context may not readily transfer to single pilot or uncrewed control regimes.

The second scope is a focus on changes concerning the human pilot in maintaining control when faced with complexity and opacity. The flight deck JCS concept defines human *actors* such as pilots, air traffic controllers, cabin crew and mission crew members, and *artefacts* such as autoflight systems, diagnostic systems and information systems. The collective actions of and interactions between actors and artefacts result in the outcome behavior of the JCS (Dekker, 2002; Hollnagel, 2005). It is well recognized that in such socio-technical systems, the entire cockpit JCS must be considered and that it would be inappropriate to regard the flight crew in isolation from other elements such as automation and other actors (e.g., air traffic controllers, ground support staff). In this light, the study reinforces the JCS

perspective on multiple occasions but maintains a scoped focus on the pilot where training and operationalization recommendations are provided. A more thorough investigation into the potential of other JCS improvements to manage complexity and opacity are considered out of scope.

The third scope is the temporal context of the studies presented. The simulator studies were performed with the knowledge, resources and background available at the time (respectively 2013, 2015 and 2018). More recent literature and studies have been sourced *a posteriori* to provide contrast with the current state of the art in the research field, but do not have a strong guiding effect on the research questions being investigated, which were relevant at the time of the studies (and still are). Results of the study, however, still very much apply at the time of writing (2026) and may prove to be even more valuable in the future contexts of HAT and RCO.

1.5 Reading guide

This dissertation consists of six chapters. The first half of the dissertation offers a retrospective investigation into existing flight deck dynamics with high levels of automation and opacity. The second half of the dissertation takes a more prospective view on how to mitigate risks associated with those dynamics. Figure 1.2 provides a visual guide to the dissertation's structure.

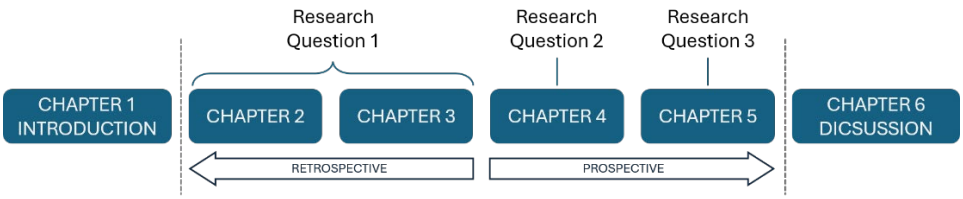


Figure 1.2. Dissertation structure overview

Chapter 2 describes an explorative study into pilot cognitive awareness (and by proxy control) of highly automated aircraft, in the context of a simulated flight scenario featuring a complex malfunction with significant opacity. The explorative study contrasts crew behavior with safety performance scores and identifies which behaviors are effective and which are to be avoided. Chapter 3 builds on the insights gained in Chapter 2 by validating whether the desirable, adaptive behaviors can be trained to induce a more effective response from flight crews when facing a complexity and opacity in a simulated flight scenario (different from the scenario in Chapter 2). This validation study introduces novel behavioral patterning techniques to allow high-resolution appraisal of behaviors in parallel with safety performance indicators.

In the second half the dissertation, Chapters 4 and 5 adopt the findings from Chapters 2 and 3 and address the challenge how to manifestation these results as realistic solutions and innovations in the existing flight deck socio-technical system. In this way, the thesis transitions from a retrospective explorative mode to the prospective application mode. Next to the empirical findings from initial chapters, Chapters 4 and 5 also incorporate perspectives from accident cases and extensive literature review. Chapter 4 addresses the challenge of harmonizing adaptive responses with compliant responses, which currently induce a problematic bifurcation in the pilot role. It proposes a new pilot model (*Airmanship 2.0*) that enables a sustainable co-existence of novel, adaptive responses with existing, compliance-based responses. This model is contrasted against six hallmark incidents and accidents that presented the flight crew with significant opacity. Chapter 5 subsequently presents a deeper appreciation of the neurocognitive foundation of sensemaking and adaptivity, proposing a training and operationalization framework. A comprehensive study into neurocognitive training of pilots is presented to further understand the potential and application of such training for the future pilot. Chapter 6 summarizes and discusses the dissertation's main findings and reflects on the broader implications and applicability of the study's results, particularly in the context of JCS design, safety investigation and future developments in the cockpit socio-technical system.

Throughout this dissertation, several terms are used interchangeably. A few are important to clarify to the reader. The term Airmanship (2.0) refers to the collection of knowledge, skills and attitudes relevant for any (future) pilot of either gender. The terms complexity, opacity and ambiguity are often used in close proximity. As these concepts lie central to the dissertation, they are defined here. Opacity is the most overarching term and is defined as an untransparent quality where information is limited, possibly conflicting, and there may be several options forward, some of which are unclear. Both systems and situations may feature opacity. Complexity and ambiguity describe different modalities of opacity. Complexity refers to the (increasing) number of systems, actors, agents, interactions, automations, etc. present in a system or operation. Complex systems and situations challenge flight crews in terms of understanding the full scope of the system, its modalities and possible failure modes. Ambiguity, on the other hand, is a quality of equivocality of options. Ambiguous systems or situations are defined as having several options which may seem equivalent, or equally uncertain. Complexity and ambiguity are independent qualities: one does not necessarily imply the other. Chapter 5 re-addresses the disambiguation between opacity, complexity and ambiguity in the context of neurocognitive sensemaking. In earlier chapters, the term opacity is used, occasionally supported by the terms complexity and ambiguity.

2

Exploring behavior in opaque operations

This chapter presents the study and findings of an initial exploration into crew behavior and performance in opaque and ambiguous flight situations. It presents a framework to specifically observe aircrew situation awareness and management of highly automated aircraft systems. The study also presents a novel technique used to stratify crew behaviors according to safety performance to identify behavioral trends that correlate with performance. It is an abbreviated publication of the larger study report published in Mohrmann (2013) and Mohrmann & Lemmers (2013).

The content of this chapter is based on:

Paper title	Investigating flight crew recovery capabilities from system failures in highly automated fourth generation aircraft
Author(s)	Mohrmann, J.F.W., Lemmers, A., & Stoop, J.
Published in	Aviation Psychology and Human Factors (2015)
DOI	https://doi.org/10.1027/2192-0923/a000079

ABSTRACT *This project aims to understand rapid crew transitions from a monitoring to a decision-making role, when asserting manual control of aircraft sub-systems. Ten crews unwittingly flew a semi-critical failure scenario in a full flight simulator, forcing several crew decision moments. Observations of automation-related (diagnostic) behavior were correlated with respective flight performance, revealing that specific competencies (related to knowledge, procedures, attitude towards automation, and teamwork with automated systems) lead to significant performance gains. More importantly, the absence of these behaviors deteriorated the crews' ability to maintain high safety margins (severely in some cases) and hence should not be underestimated in its potency to affect flight safety. These findings may form a foundation for developing and evaluating near-future innovations in training, operations and automation design, which could prove critical towards improving future accident rates.*

2.1 Introduction

Coping with growing operational demands and complexities, fourth generation [3] commercial aircraft employ an increasing amount of automation for flight guidance, planning, system management, amongst other functions. However, according to Malinge (2011), the reduction of the annual accident rate of fourth generation airlines has stagnated. Although automated systems have evolved to reduce pilot workload, recent concerns exist about their safety effectiveness. Automation introduces a paradox: providing crews with necessary operational assistance simultaneously dissociates the crew from those operations. Unfortunately, this shift in pilot tasking exhibits itself in many forms of adverse crew behavior such as automation induced complacency (Manzey, Reichenbach & Onnasch, 2012), automation bias (Mosier, Skitka, Heers & Burdick, 1998), decision making errors (Orasanu, Martin & Davison, 1998), lack of (procedural and declarative) system knowledge and/or manual control skills (Man4Gen Consortium, 2012), which are all in turn aggravated by overconfidence (Woods, 2004) and fatigue (Caldwell, 2012). These behaviors may contribute to the loss of situational awareness (SA). SA is broadly defined as the quality of awareness that an individual has of his or her (aircraft) state, system state, environment, and location (Endsley, 1995). Often these effects are precursors to a deterioration of control of the aircraft, leading to Loss of Control in Flight (LOC-I). According to the European Aviation Safety Agency (EASA, 2013), LOC-I accidents are currently the leading type of fatal accidents. Notable accident cases which exhibited tell-tale signs of this problem are Colgan

[3] This project uses the Man4Gen definition of a fourth-generation aircraft as an aircraft with full authority and full control flight (control) systems (Man4Gen Consortium, 2012).

Air Flight 3407, Air France Flight 447, Air Transat Flight 236 and, more recently, Asiana Airlines Flight 214. With the exception of Flight 3407, all aircraft involved in these accidents were fourth generation aircraft.

One initiative investigating the aforementioned automation concerns is the European project Manual Operations of Fourth Generation Airliners (Man4Gen). Man4Gen aims to understand the problems that crews face in the rapid transition from a monitoring role to an active decision-making role (Man4Gen Consortium, 2012). The project intends to develop and evaluate mitigating strategies in training, operations and system design. While Man4Gen will focus on crew response to automatic flight control system failures and flight path management, this study complements it by investigating operations with the automation of aircraft sub-systems, in particular the fuel system, focusing on lower workload, knowledge-based responses and diagnostic problem solving rather than more acute skill-based responses required for flight control issues. The fuel system is an interesting example of a highly automated system on a fourth-generation aircraft, particularly given the overall acceptance of the high level of automation in this system. Furthermore, it lends itself well for this study as its potential in both complexity and criticality may provide challenges very suitable for investigating manual control recovery. Several accidents and incidents, notably Air Transat 236, Virgin Atlantic 201 and British Airways 38, reinforce how the modern fuel system can be notoriously difficult to manage successfully in case of failures. Further elaboration on fuel systems will be presented in the methodology section. To this end, the research objective of this project is to evaluate the knowledge, awareness and problem-solving capabilities of fourth generation aircraft flight crews, during unexpected failures in highly automated [fuel management] systems. The results should indicate whether the current use of automation in the cockpit environment could be contributing to the stagnation of the reduction of the accident rate, and how. It should be noted that this paper aims to present an academic overview of this study. A more comprehensive explanation of the methodology and results can be found in Mohrmann (2013) or Mohrmann & Lemmers (2013).

2.2 Method

2.2.1 Theoretical Basis

In order to couple loss of control with a loss of awareness, a conceptual framework was developed based on the concepts of SA, as proposed by Endsley (1995), and Sensemaking (SM), as proposed by Hollnagel and Woods (2005). These two concepts are both enabled in the Man4Gen research program, and complement each other. While the SA model is effective in relating how awareness can exist in different levels, it lacks in defining how this affects a level of control. SM provides a thorough concept on different levels of control, but does not explain any base, awareness or otherwise, which would determine what

level of control is possible. Enabling both models may allude to a connection between a pilot's awareness, and their level of control. SA is divided into three successive levels (Endsley, 1995):

- Perception of the components of reality (Level 1 SA);
- Understanding these components (Level 2 SA); and
- Projection of this understanding into the near future (Level 3 SA).

The quality of each level is for a large part dependent on the quality of a lower level. SM has been defined as the cyclic process by which reality is periodically reviewed in order to update one's mental model, and distil an interaction with reality, which completes the awareness feedback loop. This so-called Contextual Control Model (COCOM) (Hollnagel & Woods, 2005), has been expanded to an Extended COCOM (ECOM) which proposes that multiple loops are stacked (Figure 2.1) and connected to create an array of control loops ranging from compensatory control to anticipatory control.

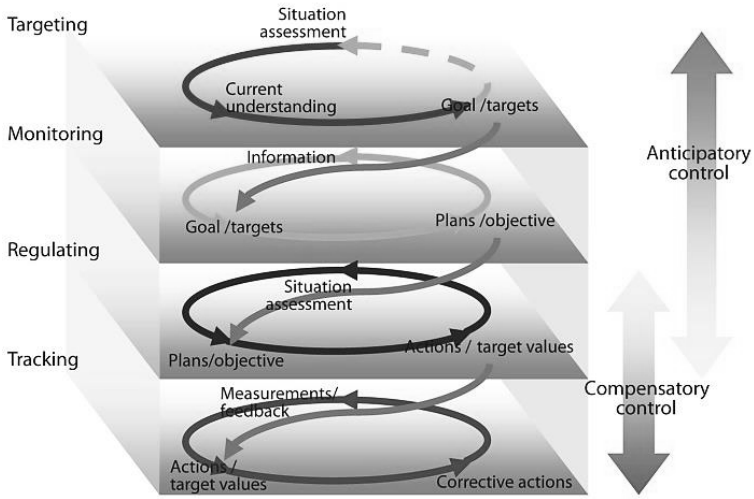


Figure 2.1. Extended Contextual Control Model (Hollnagel, 2002)

The critical link between SA and SM is that compensatory action at the tracking level of the ECOM may be coupled to a non-complex SA (i.e. perception only), whereas anticipatory control is only possible when the individual maintains excellent awareness (i.e. good projection), as illustrated at the higher ECOM levels. It follows that a fault in perception will fragment the basis of subsequent awareness levels, and that any higher level of [anticipatory] control will not interact with reality as expected. Perceiving such a cognitive mismatch between expectations and reality is experienced as the universally recognized

cognitive state of surprise, i.e. when things do not go as planned/expected (Rankin et al, 2013).

The conceptual framework enabled the quality of a decision as a central concept, used as a medium to connect the level of control of the crew to their level of awareness. At this point two important assumptions were made concerning the decisions made, enabling concepts from the theory or reasoned action (Ajzen, 1991) and the reason-action approach (Fishbein & Ajzen, 2011).

- Decisions are rational choices completely based on the information that the decision maker perceives about his reality; and
- Decisions, within the context of a specific understanding, will have a positive effect on the situation, given that this understanding matches reality.

These two premises shed some light onto the relations between understanding reality, making decisions and expected results. The four statements below further built upon these two premises and formed a layout of the final logic structure applied in this paper:

1. The level of SA is, per definition, inverse to the level of cognitive mismatch between understanding and reality;
2. The quality of a decision is inverse to the level of cognitive mismatch;
3. A particular level of control is dependent on the quality of this decision; therefore
4. The level of SA is causal to the level of control.

By this framework it was possible to connect awareness (SA-model) with a level of control (SM-model), a central premise for the conceptual framework.

2.2.2 Conceptual Framework

Automation awareness. In the context of subsystem automation, awareness was redubbed Automation Awareness (AA). The concept of AA related to areas of awareness specific to effectively operating the automation pertaining to a (fuel) sub-system. Based on the aforementioned theoretical discussion, good AA was defined as:

> *"Exhibiting sufficient knowledge, familiarity and awareness of the functioning of an automated process, in order to retain complete control of the process at any time." (author's own definition)*

Four behavioral categories were proposed as factors constituting AA: knowledge, rules, attitude and teamwork. Based on previous literature review (Man4Gen Consortium, 2012; Mohrmann & Lemmers, 2013) including the review of relevant incidents and accidents, this categorization has been devised to focus behavioral analysis and structure the

comparisons between crews. The departure point for the definition of the AA constituents was the skills, rules and knowledge (SRK) based reasoning concept proposed by Rasmussen (1983). Rule- and knowledge-based reasoning were adapted from the SRK concept as two of the four behavioral categories. Knowledge referred to both declarative and procedural knowledge of the system(s) at hand, but also of other operational details which were relevant in diagnostic or recovery steps. Knowledge is included as a behavior because it is expected to provide a critical contribution during problem diagnosis and the (re-)building of understanding. Rules referred to a pilot's familiarity with, and attitude towards procedures, checklists and other operational prescriptions and/or limitations. Rules are included because complex problem solving often involves procedures and checklists at some phase, and (mis-)management of these can have a profound effect on recovery efforts. Skill based reasoning was omitted as such reflexive motor skills were not expected to often come into play in resolving aircraft sub-system automation problems such as those studied in this experiment. The third behavioral category was pilot attitude towards automation, based on several known effects such as automation complacency, commission errors, omission errors and trust. The fourth and last category was teamwork, in consideration of the effects of social/inter-personal aspects of awareness building, notably Team and/or Shared Situational Awareness (Endsley & Jones, 1997). Especially when considering how complex diagnostics and task sharing often engage both crew members, it would be quite possible that shared awareness would be critical in effectively resolving such problems. These four discerning behavioral categories are abbreviated to the K-R-A-T decision making affectors.

Performance indicators. Upon making a decision, a crew's actions would interact with reality, and would result in some level of control, depending on if the interaction with reality was productive or counter-productive to the situation. Working from Assumption #2, it was stated that a crew's decision would always be aimed at maximizing performance (the specific performance indicators will be discussed later in this section). Therefore, it was assumed that measurable performance is indicative of a level of control. Of course, this was also dependent on the independence and measurability of performance indicators, and that enough performance variability would be present. Additionally, considering the complexities of diagnostic and recovery strategies, two different crews may enable two different strategies but still achieve the same level of overall performance. To mitigate such a nulling effect, performance was subdivided into smaller indicators (subvars) in fuel management in order to improve the fidelity of matching particular AA with particular safety performance.

Corrector factors. The final aspect of this conceptual framework was the set of three correctors which were also considered as possible unintended affectors of the quality of decision making: workload, simulator realism and fatigue. A workload that would be too high or low could introduce effects such as excessive docility or too much stress, which in

turn would affect decision making. Workload was observed by means of a RAW NASA TLX test (Hart, 2006) which uses the six workload ratings without subsequent weighting. This allows for more discretion in determining the type of workload involved. Simulator realism could affect crew immersion, where too little immersion could hypothetically result in unrepresentative behavior with respect to actual aircraft operations. The explorative nature of this study acknowledges the necessity for an ecologically valid study, and therefore proposes to at a minimum observe the pilot-perceived ecological validity. Realism was scored on a continuous scale between "no distinct similarities" and "completely accurate" for five categories: simulator hardware, simulator software, scenario setting, scenario events and an overall score. Fatigue could also severely impair proper decision making due to reduced cognitive functioning (Caldwell, 2012). Fatigue was measured by means of a discrete seven-point Samn-Perelli scale (Millar, 2012), which was chosen for its ease of administration and simplicity. If any corrector correlated with performance, then the comparability of different crews may be compromised as their behavioral bases would not be the same. Such correlations would serve to correct performance ratings before their comparison to crew AA.

The three components of the conceptual framework are combined in Figure 2.2. This framework functioned by observing pilot behavior with respect to the four AA affectors, and correlating them to their respective performance ratings (albeit corrected).

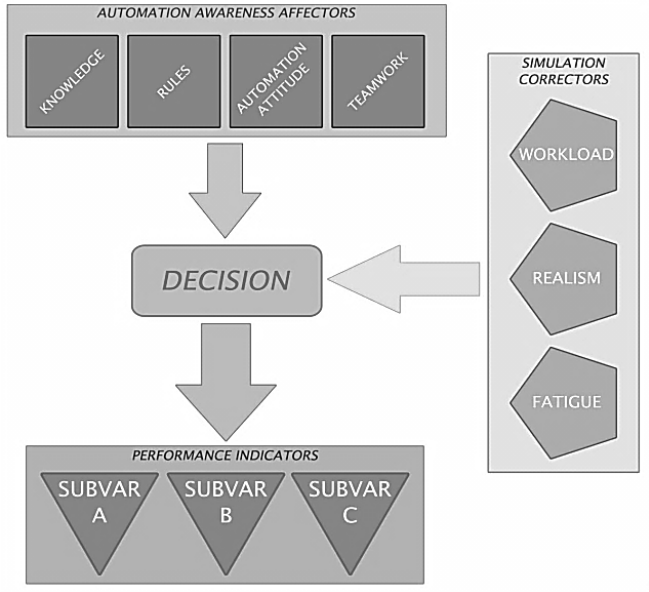


Figure 2.2. Conceptual Framework Model

2.2.3 Experiment Design

This experiment diverged from a classic control and a test group setup and focused on the observation of within-group variations of a set of representative pilot participants. Where within-group variations were limited, the study could still provide a more global insight into crew behavior and performance. This study applied a simulated flight environment in order to create a situation in which crews were forced to demonstrate their decision making with respect to fuel system automation. The choice for a fuel sub-system was made considering the criticality and complexity of such a system. Criticality was important in order to build sufficient motivation in crews to thoroughly resolve the issue. Complexity was important in order to force more expansive diagnostic procedures, which in turn allowed for the more (precise) observation of decision-making processes. The Airbus A330-200 fuel system ranked high on both factors, and featured the additional benefit of offering a larger group of type-rated participants to select from.

Scenario design. The A330 boasts very advanced automation in its fuel system. It distinguishes itself from most other fuel systems as it also fulfills non-fuel functions (such as automation center of gravity management) and automatic tank opening and isolation. This made it well suited to create a very diffuse scenario with respect to the division of control between the crew and the system computers. Essentially the scenario creates a situation in which the failure of the automatic fuel management system eventually forces the crew to assume manual control of the fuel system. The scenario also introduces a leak in the fuel system, which is difficult to diagnose yet critical enough to warrant resolution. By this the scenario challenges crews in their capacity to correctly diagnose the problem, recognize what action on their behalf is required and execute an effective course of action. The scenario was set in a trans-Atlantic Extended Twin-engine Operational Performance Standards (ETOPS) flight from Europe to the Caribbean region, where the routing was very southerly, crossing the Azores and remaining equidistant between the African west coast and the U.S. east coast. The reason for this routing was a tropical weather cell in the northern Caribbean region. The scenario would start about two-thirds into the cruise phase, approximately 25 minutes prior to the flight plan's critical fuel point (beyond that point the fuel plan ensured 15 minutes of fuel upon arrival, given an emergency descent to FL100 and continuing with one engine inoperative - an ETOPS flight planning condition). Furthermore, the preflight paper briefing listed two components which would be unusable during the flight (legal and according to the Master Minimum Equipment List). One was the Fuel Control and Management Computer (FCMC) No. 2 (of 2), which was important in relation to the FCMC 1+2 FAULT (loss of automatic fuel management). The other unusable component was a hydraulic accumulator fault, which had no further consequences and only served to distract the crew from the fuel system. During the beginning of the scenario several failures and events steered crew

expectations by priming their mind-set, which set the stage for misleading, confusing or surprising the crew which creates the diffuse setting during which the fuel leak is subtly introduced. This leak was further cloaked by faulted (fuel management) automation, which forced the crew to assume manual control of the fuel system. As the leak was central to the scenario, the leak rate was finely adjusted in order to provide the crew with sufficient time/fuel to reach several diversion fields, but to also be large enough in order to be noticed and evoke a sense of urgency. The leak rate of 1.5 kg/s was tuned to finely balance realism and challenge required to fully engage participants in their decision making process. The scenario was developed and validated together with operational experts on the A330 – both flight crew and system experts. All sessions were identical up to a particular Event 7, which is when all system anomalies and failures stop evolving, after which crews were able to diverge in their recovery attempts. Therefore, crew-specific performance is only measured after Event 7. These performance variables collectively represent the safety margin that the crew can maintain (i.e. how much fuel they have on arrival). These variables are further explained in the Results section. Crew behavior was monitored throughout the entire scenario, as behavior prior to Event 7 could also have had an impact on their course of action after Event 7. A detailed account of the scenario schema, including Event 7, can be found in (Mohrmann & Lemmers, 2013).

Resources and planning. The experiments were hosted at the National Aerospace Laboratory (NLR), in the Generic Research Aircraft Cockpit Environment (GRACE) simulator. The generic cockpit could be modified to accurately represent several common cockpit environments, including the A330. The simulator also featured functionally sufficient automation and FMS features, and radio contact with Air Traffic Control (ATC) in the simulator control room. The observer was seated inside the cockpit aft of the two pilots, and also played the occasional role of purser/cabin crew. Confederate air traffic controller roles were fulfilled by other A330 pilots familiar with mid-Atlantic radio communication, although most of the ATC responses were anticipated and scripted ahead of time. Participants were invited for an "observation study", and not informed about the true nature of the scenario. Participants qualified if they were active duty A330 captains, first officers (F/O) and second officers (S/O). As the scenario would be a cruise situation, the following combinations were considered valid:

- Captain & first officer
- Captain & second officer
- Two first officers
- First officer & second officer

There was no further discrimination according to experience or flight hours. The only demographic goal of this project was to obtain a random set of the above valid cruise crew combinations. Variations in crew composition are the result of scheduling, and were not explicitly targeted. Participants were 20 active duty A330 pilots (four captains, 11 first officers and five second officers), forming 10 crews. The experiments took place during one week to limit possible cross-contamination of subject experience. Simulations were held during normal working hours to mitigate possible fatigue effects. Upon arrival, participants were familiarized with the simulator after which they were permitted to self-brief with a representative yet modified dispatch briefing document, and were informed they would start around two-thirds into the cruise phase. The experiment scenario duration varied between 50 and 75 minutes depending on a crew's decision making speed. Each scenario ended when the crews practically exhausted their diagnostics, and communicated a diversion plan.

Observations. Several different modes of information were recorded. The simulator recorded important flight and fuel system parameters (e.g. location, fuel tank quantities, failures). The simulator also facilitated audio and video recording of the pilots and displays. The observer recorded behavior, decisions and other notable occurrences during the flight. All sessions are observed by the same observer in order to mitigate inter-rater disagreement. Participants filled in a post-hoc self-reflection form using both visual analog scales and open text responses to guage corrector variables, the individual pilot's attitude towards automation, with six questions pertaining to the respondent's feelings of trust, dependence, familiarity, etc. with respect to on-board automation in general. Lastly, the observer engaged both crew members together in a cognitive interview session, during which several important or unclear events were discussed, such that pilot rationale could better supplement observed behaviors. Further details about forms and the data results in Section 2.3 may be found in Mohrmann (2013) and Mohrmann & Lemmers (2013).

2.3 Results

2.3.1 Performance Indicators

Essentially, the most fundamental indication of good performance could be represented by the safety margin created by pilots. Given the nature of the failures, maximizing fuel on arrival (FOA) would be a heuristic for maximizing the safety margin during this emergency. The authors acknowledge that in other scenarios (e.g. with more airfields in the near vicinity) landing with more or excessive fuel does not necessarily contribute to safety. However, considering the fact that this fuel emergency is occurring at a low, even critical, fuel state, any extra fuel margin does contribute to ensuring flight safety. Table 2.1 depicts the fuel upon arrival. For comparison, no crew action would have the aircraft arrive at an enroute diversion field with an estimated 3935 kg of fuel. For another

comparison, the planned arrival fuel for the original destination is 5700 kg. In Table 2.1, as well as the other performance tables, the second row indicates which crew combination is associated with each performance, which indicates whether particular crew combinations are more successful than others. These crew combinations will be discussed in the "Effect of Crew Combination" sub-section.

Table 2.1. Fuel on Arrival values ranked from best to worst

Best performance				**Fuel on Arrival (kg)**					**Worst performance**
4470.5	4020.2	3886.5	3688.1	3542.0	2799.1	2520.5	512.0	409.6	-500.3*
F/O-S/O	Capt-S/O	F/O-S/O	Capt-S/O	F/O-S/O	F/O-F/O	F/O-F/O	Capt-F/O	F/O-F/O	Capt-F/O

**This negative value indicates that the crew had insufficient fuel to reach the airfield.*

However, a single fundamental indicator was not sufficient in comparing two different crew strategies leading to the same FOA. Therefore this study introduced three sub-performance variables (subvars): amount of fuel leaked, fuel burn efficiency (kg/NM) and the distance flown. The combination of these variables fully represented all the possible ways to reduce the fuel quantity on-board, which was directly causative to the safety margin achieved. At this point it is important to note that the scenario differed slightly between crews, due to certain events being triggered by (variable) crew actions. All events up to Event 7 were effectively priming events and did not warrant any differences in crew diversion actions. Therefore all performance (sub) variables were only evaluated from Event 7 onwards. Additionally, because the crews did not fly the aircraft until landing, their intentions for the remaining 1-1.5 hours (clearly verbalized before the end of each session) were used to extrapolate performance according to Airbus A330 performance specifications.

Subvar A: Amount of fuel leaked. This variable indicated how much fuel was lost through the leak after Event 7 until landing. For many crews they (nearly) exhausted all fuel which was available to the leak (which is approximately half of all fuel on-board). The fuel leaked was calculated by finding the difference between the total reduction of fuel (between Event 7 and landing), and the amount of fuel burned. This subvar was also corrected for the fact that not all fuel can be leaked, and that the "leakable" amount of fuel varies with Event 7 timing. Subvar A provided an objective score of how well the crew was able to minimize fuel loss through the leak. Because the leaking tank would always be empty well before arrival, difference in diversion distances and airspeeds did not affect this value. Table 2.2 shows subvar A scores.

Table 2.2. Percent of leaked fuel which leaked ranked from best to worst

Best performance		Percent of leakable fuel actually leaked						Worst performance	
41%	57%	60%	61%	63%	68%	68%	69%	73%	94%
Capt-S/O	Capt-F/O	F/O-S/O	F/O-F/O	F/O-S/O	Capt-F/O	F/O-S/O	F/O-F/O	Capt-S/O	F/O-F/O

Subvar B: Fuel efficiency. This variable indicated how efficiently the crew configured their aircraft. This depended on a number of factors including the number of engines running, the altitude and airspeed. Some crews indicated actions that would affect the efficiency at some point well beyond the duration of the simulation. This information was used to correct the extrapolated efficiency. Using the distance flown since Event 7 plus the extrapolated distance to the diversion airport, this was combined with the total amount of fuel burned between Event 7 and landing, which together determined the amount of fuel used by the engine(s) per nautical mile (NM). Airbus performance data was used to extrapolate fuel burn after the end of the simulation depending on the chosen cruise altitude and engine configurations. Table 2.3 displays the subvar B performance rankings.

Table 2.3. Fuel efficiency ranked from best to worst

Best performance		Fuel efficiency (kg/NM)						Worst performance	
7.56	7.93	9.03	9.41	10.11	11.18	12.21	12.48	12.68	13.74
Capt-S/O	F/O-S/O	F/O-F/O	F/O-S/O	F/O-F/O	F/O-S/O	Capt-F/O	Capt-S/O	F/O-F/O	Capt-F/O

Subvar C: Distance overflown. This last sub-variable indicated the quality of a crew's diversion decision. Ideally, a good decision was made without undue delay, and to the nearest airfield. Upon Event 7, a crew would have the minimum information available to make a diversion call, even without a diagnosis. This variable calculated the nearest valid diversion field at Event 7, and compared this reference distance to the actual distance flown after Event 7 until landing. This was expressed as a percentage in which 20% overflown indicated that the crew flew 1.2 times the minimum achievable distance from Event 7. Both the time to make a decision as well as choosing the nearest most suitable airport were represented by this statistic. Table 2.4 displays the subvar C performance rankings. Despite the fact that this measure may suggest promoting fast and hasted decision making (which is not the authors' intentions), it attempts to provide a clear perspective on differences in how crews take into account their diversion options and decision making, irrespective of other,

also important, considerations such as failure diagnostics and continued flight operations (represented by the other two subvars).

Table 2.4. Percent distance overflown ranked from best to worst

Best performance		**Distance overflown (% of nominal Event 7 distance)**						**Worst performance**	
6%	13%	14%	18%	31%	35%	37%	39%	40%	49%
F/O-F/O	Capt-S/O	F/O-S/O	Capt-F/O	Capt-F/O	F/O-F/O	F/O-S/O	F/O-S/O	F/O-F/O	Capt-S/O

Regression analysis. The three subvars were constituents of the total FOA. However, the fundamental safety margin of FOA could not be used as it was not easily corrected for Event 7 timings. In order to correct for Event 7 timings, the FOA variable was replaced by a heuristic variable based on the amount of fuel used (FU: both leaked and burned) after Event 7. The FOA and FU/NM variables were regressed to each other and aligned very closely ($p<0.001$, $R^2=0.96$). Subsequently, the three sub variables were regressed to the FU/NM heuristic variable using an Ordinary Least Squares (OLS) stepwise regression by backward elimination, and also exhibited a strong (explanatory) correlation ($p<0.001$, $R^2=0.946$).

2.3.2 Correction Factors

The three corrector factors were regressed to the four performance variables (three subvars and FU/NM). All values were averaged per crew. All 12 regressions use crew-averaged values and followed stepwise regression by backward elimination of corrector sub-indices, with performance indices as the dependent variables. The only two regressions which exhibited marginally usable R^2 (0.67, 0.64) values were the workload and realism relations with subvar C. However, these two R^2 values do not indicate sufficient predictability to warrant adapting the performance indicators with them. These R^2 values do provide an indication, albeit rough, that the performance indicators are representative of AA only. Of course, it must be noted that such correlations must be used with caution, in light of the small sample size. However, the exploratory nature of this study warrants at least a preliminary analysis of potential corrector effects. Future studies with larger sample sizes may be able to fortify such correlation studies. Even without clear regression results, average ratings could still shed light on the quality and suitability of the experimental setup for evaluation purposes. The results indicated that a mild workload was present, with a notable increase in mental demand. The realism scores showed moderate scores for hardware and software ratings, but much better ratings for the scenario. The fatigue ratings indicated that,

according to the Samn-Perelli scale, on average, participants felt "OK, somewhat fresh" after the simulation session.

2.3.3 Automation Awareness

Behavioral observations which constituted AA were sourced from live observer notes, audio/video recordings and interview notes, with the addition of an automation attitude questionnaire collected via the self-reflection form. This questionnaire explored attitude to automation in general, not specifically the fuel system. Regression of the questionnaire results with the performance variables did not reveal any significant relations, however it was interesting to note that on average pilots felt they were in control and had good oversight, but were less confident about knowing the role and necessity of automation in controlling the aircraft.

Besides the questionnaire, the majority of the observations underwent processing prior to being merged with performance data. Observations were refined for clarity and subsequently initially classified according to the four AA affectors, and secondly classified according to their positive or negative contribution to the three performance subvars, creating 24 categories in total. It should be noted that all observations were processed without knowledge of that session's performance. This was to prevent a performance indication bias during categorization.

2.3.4 Effect of Crew Combination

Reviewing the crew combinations indicated above per performance variable, these combinations seem to shift in performance ranking depending on the performance variable considered. In order to analyze any general trends in the effects of crew composition, Figure 2.3 depicts a count of each type of crew rank (captain, F/O, S/O) per performance ranking (from best to worst). Working from the premise that a higher flight rank indicates more experience in the A330, this figure will indicate if there is more experienced pilots also perform better.

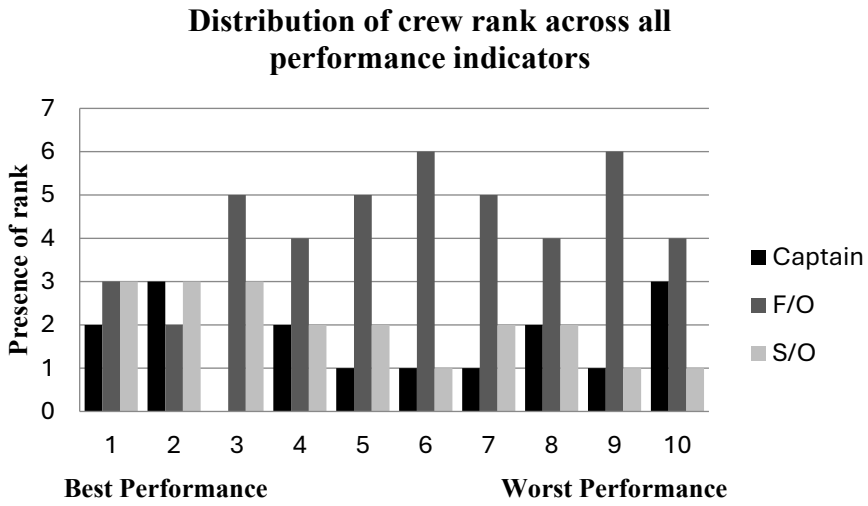


Figure 2.3. Distribution of crew rank across all performance ratings

What is noteworthy from Figure 2.3 is that captains are increasingly present in both very good and very poor performing crews. While F/O's have a mild peak in the middle, S/O's are notably more present in the better performing crews. As the performance of a crew may not only be attributable to their total experience but also to the combination itself, Figure 2.4 depicts the presence of the four possible (cruise) crew combinations according to the same performance ranking.

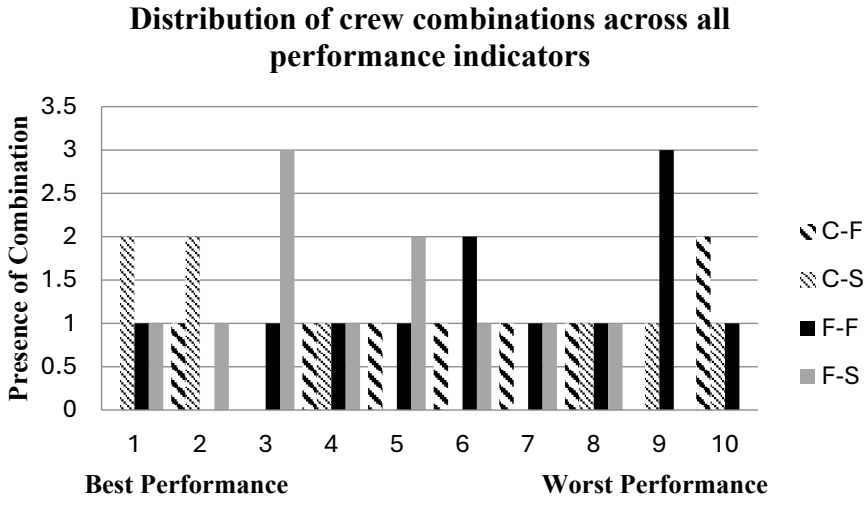


Figure 2.4. Distribution of crew combinations across all performance indicators.

Figure 2.4 indicates that most crew combinations seem to be fairly well distributed across performance, with a small increase in captain-S/O crews as better performers, and more F/O-F/O crews in the poorer half of performers. In light of Figure 2.3 and Figure 2.4, and noting the small sample size of 10 crews, the authors conclude that variations in crew combinations are not profoundly linked with variations in performance, and therefore will treat all crews as equal in this study. Perhaps a larger sample size would indicate if such crew variations do have an effect, but this cannot be confirmed at this point.

2.3.5 Combining Behavior with Performance

When considering crew variations with performance, all crews throughout all rankings are used, providing a wide stratification of the dataset. The reason for this is that crews may perform similarly with different behavior aspects being observed. In other words, not all behavioral aspects are as clear for every crew, therefore some behavioral distinction may also be found in performance variations amongst mid-range performers. Within each category, the ten sessions were ranked from best to worst pertaining to that specific performance subvar. This was the first time that performance data and behaviors were matched together. By aligning the results, it became clear that good performers tended to have more positive and less negative comments, and vice versa, which is a mild confirmation of consistent observation. At this point a process called Recursive Abstraction (RA) was started, which essentially performed several subsequent summarizing steps in order to combine observations into increasingly more general trends. For this, the "Ladder" technique was devised, which was a progressive stepwise analysis matching changes in performance with changes in behavior. The Ladder technique began with the poorest performer's positive remarks, and then observed which behavior the second-to-last poorest performer exhibited, inferring that this change in behavior was causal to the improvement in ranking. The same was done for the negative comments, but worked in the opposite direction. The Ladder technique also employed a second phase in which the actual quantitative changes in performance were used to indicate which behavioral changes were more important (i.e. resulted in greater changes in performance). As a validation step, these behavioral changes were compared to the pilot responses to open questions about their experiences of surprise, unclarity and/or uncertainty. Only a few minor adjustments were required; on the whole the pilot and observer comments aligned well. The resulting 24 small cohesive descriptions of suspected causal changes in behaviors underwent RA again and were integrated across all three performance variables, resulting in eight conclusive paragraphs describing how each AA category either positively or negatively affected performance in general. provides a very brief overview of these eight paragraphs. Behaviors with a positive effect are entitled "behavior to promote", and behaviors with a negative effect are entitled "behavior to avoid".

Table 2.5. Summary of behavioral results

	Knowledge	Rules
Behavior to promote	Extensive knowledge of relevant technical details Work effectively with a mental model of the problem Rules-of-thumb being recalled Good mental model of the current aircraft state	Assume worst case scenario Complete understanding of structure/diagnostics Understanding procedure limitations Plan of action/read through
Behavior to avoid	General lack of understanding of system operation Poor mathematics and calculations Unfamiliarity with various failure cases Not suspecting unlikely occurrences	Distrust procedures Poor awareness of suitable checklists Do not read through the checklist Do not go beyond the checklist
	Automation attitude	Teamwork
Behavior to promote	Familiar with non-normal ops increases automation acceptance Level of wariness, noting other information sources Increase in automation acceptance increases performance	Generate options, have a plan Active cross-checking Setting priorities Task sharing with multiple players
Behavior to avoid	Lack of deeper interest in automation functioning Attempt multiple resets Assume problem will fix itself Increased level of distrust; assume computer is always the culprit	Level of distrust in others Increased reliance on others Poor idea sharing Too explicit task splitting

2.4 Discussion

The conclusions drawn in this study based their validity and usefulness for a large part on the realism that the scenario obtained. The realism evaluation scores indicated that the scenario and its events rank as very realistic, but the simulator hardware and software less so. The low simulator rating was for the most part based on the warming-up scenario where the manual flight control model was not suitably calibrated for the A330. The experimental scenario only employed auto flight features which functioned normally, hence this study concluded that the low score did not weigh as much as it may seem. The high scenario scores were important in confirming crew immersion into the scenario. With the suitable exception

of mental workload, results indicated a low overall workload during the scenario, which was sought after to avoid contaminating action-outcome relations with non-cognitive workload. In addition, the low correlation between correctors and performance scores indicated that the scenario was more or less equally challenging and realistic for all crews, ratifying the comparability between crews. To conclude, this scenario was arguably a sufficient basis to draw conclusions about crew behavior in real life.

The second challenge in this study was to obtain objective ratings of performance. The scenario design intended to limit the number of contingencies without making it too obvious which options were possible. As detailed in the Methodology section, the fact that the fuel system was severely impaired, that the scenario develops near a critical fuel point and that the fuel margins upon arrival are comparatively low, it can be argued that reduced flight endurance represented the largest flight risk at that time, and the best possible actions were arguably those which increased the fuel margin upon arrival. Subsequent use of sub-variable performances provided the detail in explanatory performance indicators in order to differentiate and grade different crew strategies leading to the same fuel upon arrival. In this way the scenario succeeded in creating an objective basis (in this case safety-related) for condemning or promoting certain behavior.

The third challenge in the setup of this study was to facilitate fair grading of crew behavior. Inherent to the fact of humans making observations, bias and subjectivity are never completely removed. However, this methodology employed five techniques to mitigate subjectivity:

1. The sheer number of initial observations dampened any observer opinions in specific observations;
2. All observations were verified between three different sources (notes, recordings, interviews);
3. Early in the process, observations were stripped of metadata, and were always treated separately from performance data (they were analyzed before the performance data);
4. The *Ladder technique* performed an additional validation with pilot self-reflection remarks not used before;
5. Recursive abstraction enabled many (24) rankings based on objective performance, such that biases not already neutralized were suppressed by the generalization process;

In addition to this, throughout the RA process, comments seemed to line up very well as every stage, and positive and negative comments would often show complementary behaviors. Furthermore, the fact that the number of positive and negative remarks correlated

with performance as expected (i.e. good performers have more positive and less negative remarks, and vice versa), indicated consistency in the observation process.

The above provided evidence of a sound basis upon which conclusions may be made. An additional support to the (ecological) validity of the findings can be found as observed behavior which echoes previously documented phenomena. Automation complacency (Manzey, 2012) was demonstrated by crews waiting for further warnings, expecting malfunctioning automation to still support their decision making and hesitating to take over from the automated system. Automation bias appeared more pronounced in the form of "checklist bias", where crews blindly assumed that the procedures would deliver clarity and resolve the issue, when in many cases such overreliance only made the complex failure difficult to understand due to changing aircraft configurations. It should be noted that both automation and checklist biases were more prominent in pilots with less flight experience (in the A330). Crews which exhibited a better understanding of the way the systems work, in some cases extensive technical knowledge, as well as familiarity with the limitations of the procedures, noted these assets to be very beneficial in deciding which checklists to apply and setting an effective course-of-action. Both assessment and course-of-action decision making errors (Orasanu et al, 1998) were identified in, for example, misdiagnosing slowly changing system states (assessment error) and poor critical evaluation of the outcome of a checklist (a course-of-action error). In some sessions, certain cognitive challenges resulted in stress and tunneling which were expressed by cursing, holding on to the incorrect memory of a display and locking into an incorrect diagnosis. Crews performed better when they cross-checked actions with each other, and were able to check their expectations for system responses with observations from other information sources. Crew confidence worked well in some cases with experienced pilots, but overconfidence in those less experienced often backfired into worsening surprise and confusion after noticing unexpected system responses upon checklist directives. This study also reinforced results from another study (De Brito & Boy, 1999), in which survey results indicated that crew were in many cases unsatisfied with the current complexity and user-unfriendliness of checklist procedures; although this could arguably be the outcome of a systemic lack of knowledge about the systems concerned in the checklist, which was demonstrated in this study. The few crews with a more extensive understanding of how a fuel system operates (in general terms, not specifically the A330) were able to translate this knowledge to the checklist procedures of the A330, and arguably apply procedures more effectively.

In conclusion, this study demonstrated that it is possible to effectively assess crew behavior in an ecological setting, using objective performance indicators. Based on the results of this study, four statements were be made in lieu of the research objective:

1. There were clear differences in crew behavior with automation, despite equal type-qualifications.
2. Variations in AA have demonstrably resulted in a significant variance in safety margins – poor AA directly results in poorer decisions which, in this scenario, led to a reduction in safety margins.
3. Negative AA behaviors detailed in literature have been demonstrated to exist in an ecological context and also result in degraded performance, as proposed.
4. Variations in rank or crew composition did not correlate with variations in AA.

This study concluded that variations in crew interaction with automated systems should not be underestimated in their potency to affect flight safety. With aviation design and operational trends bolstering automation capabilities, minimizing training standards and simplifying the role of the pilot, the effects of AA on safety may become more pronounced without due consideration for the unintended side effects. In light of a stagnating accident rate, the methodology and results of this study may help identify which AA mechanisms and behaviors should be prevented but also, more importantly, identify positive examples of AA mechanism and behaviors which should be promoted. Further studies can use these analyses to direct new training, Human Machine Interfaces (HMI) and operations design initiatives to effectively improve the crew-automation interaction system. The Man4Gen project is already paving this path upon their experimental results.

Although this project has succeeded in identifying behaviors which impact performance and safety, it does not yet explain how these (variations in) behaviors come to be in the first place. An attempt at correlating behavior with specific attributes such as flight experience, leadership skills, technical knowledge, etc. may help in further defining which competencies could be spearheaded in training and selection programs. An example of such an attempt is Project Samurai (Gorter & Jaeger, 2014), which attempted to investigate whether biofeedback training could improve manual flight control by accelerating psychological and somatic recovery after stressful events. Finding the source of behaviors may be critical in defining the effective solution.

DISCLAIMER *This research was co-funded by the National Aerospace Laboratory (NLR) and as part of the European Commission FP7 2012 Aeronautics and Air Transport program under EC contract ACP2-GA-2012-314765-Man4Gen. The views and opinions expressed in this paper are those of the authors and are not intended to represent the position and opinions of the Man4Gen consortium and/or any of the individual partner organization.*

3

Evaluating a new strategy for opaque operations

This chapter present the results of a study seeking to validate the possibly of training desired behaviors identified in Chapter 2. A novel problem-solving strategy and related Risk Information System display were evaluated in a simulated scenario. Findings confirm the desirability of certain behaviors in the context of ambiguity and complexity, as well as the necessity for more robust training to augment behavior.

The content of this chapter is based on:

Paper title	Training sensemaking for the modern flight deck: A new perspective
Author(s)	Mohrmann, J.F.W. & Field, J.N.
Published in	Pre-print (engrXiv, 2026)
DOI	https://doi.org/10.31224/7409

ABSTRACT *This chapter evaluates a novel problem-solving strategy for flight crews experiencing complex, ambiguous and opaque situations in modern, highly automated (fourth generation) aircraft. This strategy has been developed and evaluated in the EU research project Man4Gen in 2015, and provides a structured approach to managing time, reducing situational opacity and managing complex contingencies. The strategy is based on initial exploratory research in previous Man4Gen simulation studies. An experiment was performed with 14 flight crews type-rated for either the Airbus A330 or A320 from various airlines, presenting flight crews an ambiguous and opaque situation in a carefully refined simulated scenario featuring a complex energy management issue combined with a diversion in poor weather. Crew performance was evaluated on 30 distinct challenges and risks that could present themselves throughout the scenario. These performance metrics were compared to crew behavioral indicators concerning the use of the novel strategy and supporting risk information display innovations. The experiment features three sets of crews: baseline crews (N=3 crews) did not receive training nor displays, one set of crews (N=7 crew) received only training in the strategy and were provided with a quick reference card, and a third set of crews (N=4 crews) received training in the strategy as well as access to the supplementary Risk Information System (RIS) designed to work in tandem with the strategy. This chapter evaluates both the strategy training and RIS impact on behaviors (through a between-group analysis), as well as the relationships between behaviors and safety performance, by exploring correlations in a within-group analysis of the entire population (N=14). This allowed for a more robust experiment setup, with less sensitivity to random effects in small group sizes. Results on training effectiveness indicate that the training alone did not induce the desired behaviors as hypothesized (and sometimes the opposite). The RIS was effective in increasing behaviors related to contingency management. The second analysis confirmed that performance increase correlates with several hypothesized behaviors, but results were not powerful enough above post False Discovery Rate p-adjustments. Behaviors related to increased performance included efficient short-term time management before problem solving and increased time spent managing uncertainties and contingencies in a sensemaking cycle. Although the experiment featured several limitations and can be improved upon in terms of training design and possibly deploy a longitudinal study design, it presents a concrete first attempt to both operationalize and visualize resilience theory by engaging human sensemaking abilities in opaque and ambiguous situations.*

3.1 Introduction

The past two decades have shown a shift in aviation safety and accident causation. While the safety performance of the aviation transport system is unprecedented, accidents are increasingly characterized by difficulty of flight crews asserting manual control, maintaining situation awareness and demonstrating effective decision making (Sarter & Woods; 1994; Saurin & Carim, 2012; IATA, 2014b; Man4Gen Consortium, 2015; Mohrmann et al., 2015; Stoop & Van Kleef, 2015; Wolter et al., 2015; Strauch, 2017; Kharoufah et al., 2018; Proctor & Van Zandt, 2018; Kelly & Efthymiou, 2019; Prinzel et al., 2024; Gago et al., 2025). This is in part attributable to the increased reliability of systems which erode crew familiarity with system intricacies and malfunctions, in addition to an increased focus on (procedural) compliance in crew training and operations. While the emphasis on procedural compliance is understandable, it may also be ossifying crew competencies to act effectively in non-standard situations (Landry, 2009; Mohrmann et al., 2015; Rankin et al., 2016; Abbot, 2017; Niedermeier et al., 2018; Boy, 2020; Clark & Wilson, 2024), expanding the chasm of control between crew and systems.

To better understand the nature of this shift in accident causation and how it can be mitigated, the EU Man4Gen project ("Manual Operations of Fourth Generation Airliners", 2012-2016) investigated human-machine interactions in modern, highly automated fourth generation aircraft (Man4Gen Consortium, 2015). The need for such an exploration is also reiterated by DeSalvo & Fogarty (2016) who underscore, from a system verification and validation point of view, the necessity for "aircraft-level operation" validation of systems, instead of only single system-level and functional-level validation. Such limited scopes of validation will too often lead to a myopic set of assumptions of system functionality, when the "cockpit operational ecosystem" may suffer from clutter, poor integration and (most notably) ineffective human-machine interactions. Furthermore, DeSalvo & Fogarty (2016) indicate their concern with the assumptions about the real-world accuracy of system simulation models, as well as the safety assessment of changes in highly integrated, complex operational systems. In the context of such complex systems, verification and validation methods must shift from isolated system validation, to integrated and operationally representative validation of the collective system, including human actors. It is only through such holistic approaches that sufficient insight can be provided into cascading effects and dynamic relations within increasingly coupled system architectures. Such a perspective on system validation also meshes well with the Joint Cognitive System (JCS) perspective (Hollnagel, 2005). The necessity for such a holistic approach was further underlined in the US Senate's Committee Investigation Report concerning the certification efforts of the Boeing 737MAX (Senate Committee, 2020), in which long-held human factors

assumptions were left unchallenged, despite growing system complexity in primary flight control systems.

This chapter seeks to validate whether a novel problem-solving strategy designed toward managing complex and ambiguous situations can induce effective crew behaviors and improve flight safety in such opaque and ambiguous situations. The strategy and associated display concepts were developed, implemented and tested by the Netherlands Aerospace Centre (NLR) and German Aerospace Centre (DLR) in the Man4Gen project. Validating this strategy, given the complexity, human variability and scenario flexibility, calls for a novel, holistic and data-driven approach to compare flight crews in behavior and flight safety performance indicators. This chapter will therefore evaluate the effectiveness of the proposed resolution strategy as well as the application of new behavioral and performance analysis methods suited for complex and unpredictable research contexts.

Section 3.2 will provide research background of the Man4Gen project, introducing both the proposed strategy and risk display support solutions and validation approach. Section 3.3 describes the research method, evaluation context (scenario) and behavioral and performance measures. Sections 3.4 and 3.5 present study results and conclusions, and Section 3.6 discusses the larger operational and research implications for human performance in complex and ambiguous systems and operations.

3.2 Man4gen project background

3.2.1 Summary of the Man4Gen strategy design

The EU research project Man4Gen explored the difficulties that flight crews experience in asserting manual control, both physically and cognitively, in modern fourth generation aircraft that feature high levels of automation. Based on the initial Man4Gen exploratory simulation experiments studying these control challenges (Mohrmann et al., 2015; Rankin et al., 2016; Niedermaier et al., 2018), the NLR and DLR teams in the Man4Gen project articulated training, operational and cockpit display innovations to support flight crew recovery of situation awareness and cognitive control (i.e., understanding the state of the aircraft and its systems, recovery options and mitigation options) in the face of opaque and ambiguous situations that modern, complex aircraft may provide (Field et al., 2017; Buch et al., 2017). These innovations distinguished themselves from existing problem-solving guidance and methods in three distinct areas.

The first is a differentiation from common industry standard approaches such as T-DODAR, FORDEC and DESIDE [4] (Banks et al., 2020) which emphasize "identify the

[4] T-DODAR: Time, Diagnose, Options, Decide, Assign, and Review

failure" or "diagnose" and proceed linearly to flight continuation planning. While such a linear process suffices when failures are singular, distinct and recognizable, complex failures in modern airlines do not always present themselves in such a transparent way and may feature significant opacity and ambiguity. Rather, previous Man4Gen simulator studies (including the one presented in Chapter 2) indicate that a *cyclic* learning process which alternates between sensemaking and execution was observed as a more effective approach to understand and resolve complex, multilateral and/or diffuse non-normal situations.

Second, existing problem-solving strategies are execution-focused and feature little guidance toward stabilizing the situation or buying time (which the exception of *Time*-DODAR). In the exploratory Man4Gen studies including the one presented in Chapter 2, poor-performing crews were observed to experience high temporal pressure and often a strong drive to land the aircraft as soon as possible. While not a bad objective in and of itself, when returning to land *becomes* the short-term plan, additional complexity and risk is introduced when requisite sensemaking occurs simultaneously with high workload flight phases that also feature system reconfigurations (i.e., approach and landing). As such, effectively managing time and "taking a moment" for sensemaking was shown to positively contribute to a crew's ability to manage these complex situations. The necessity for such "self-recovery" is echoed in findings from more recent studies in startle and surprise management techniques (Field et al., 2018; Landman et al., 2020).

Third, the interaction between crew members and systems within the flight deck socio-technical system may also feature innovations that can make (or break) such sensemaking for opacity. Existing failure management systems such as the Airbus Electronic Centralized Aircraft Monitor (ECAM) and the Boeing Engine Indicating and Crew Alerting System (EICAS) are useful interfaces to present a system's current state but are poor substitutes for diagnostics (for which they are often used). By explicitly decoupling system state information and executive guidance in system presentations, cognitive space can be created for the flight crew to aggregate a variety of contextual and state information and engage in more effective Threat and Error Management (TEM) techniques (Helmreich et al., 1999). Such a revision in systems allows a broader transition in the cockpit socio-technical system to support sensemaking.

The Man4Gen project developed a novel problem-solving strategy and associated Risk Information System (RIS) to evaluate whether flight crews can be trained and guided towards effective problem solving in opaque and ambiguous situations. This study focuses

FORDEC: Facts, Options, Risks and benefits, Decide, Execute, Check

DESIDE: Detect, Estimate, Set safety objectives, Identify, Do, Evaluate

on the evaluation of the effectiveness of this problem-solving strategy and associated systems in the context of a fourth-generation aircraft scenario featuring opacity and ambiguity. Additional details in the design and development of this strategy may be found in a sister publication (Field et al., 2017). The underlying philosophy behind this strategy consists of three key principles, in the following order:

1. Time Management (TM),
2. Uncertainty Management (UM), and
3. Contingency Management (CM)

These principles were operationalized in six phases, with each phase actively engaging the crew cognitively, in contrast to merely prescribing actions as traditional checklists and procedures often do. The phases are listed in Table 3.1. Phases 1, 2 and 3 were intended to manage time criticality; Phases 4 and 5 manage uncertainty; Phase 6 supports planning for contingencies and changes.

Table 3.1. Overview of strategy phases and underlying philosophy principles

Principle	**Phase No.**	**Phase**
Time Management (TM)	1	Stabilize flight path
	2	Mitigate immediate threats
	3	Short-term planning
Uncertainty Management (UM)	4	Identify situation
	5	Perform appropriate actions
Contingency Management (CM)	6	Long-term planning

Crews are expected to sequence the strategy from Phase 1 to Phase 6, and then cycle between Phases 4, 5 and 6 to engage in a continuous sensemaking process. This is in line with the recursive sensemaking concept proposed by Hollnagel (2005). As complex, ambiguous, opaque situations are often unclear when they first present, they clarify as crews repeatedly hypothesize, experiment and learn about the situation. The initial Phases of the strategy have their primary purpose to create time and space for sensemaking to occur (i.e., providing a temporary, stable situation and removing temporal stress on the crew). These initial phases are critical yet intended to be performed usually only once. Crews trained in following this strategy are provided preparation material, classroom training and a quick reference card Appendix A) as a mental aid containing the most important considerations for each phase. More information about the strategy development and design process can be found in (Field et al., 2017).

3.2.2 Extending the strategy with the Man4Gen Risk Information System

Alongside the development of the strategy, the Man4Gen project also investigated a new cockpit display concept to support pilots in regaining manual control. This concept described in detail by Buch et al. (2017) is referred to as the Risk Information System (RIS) which aims to provide flight crews more context-specific risk information, to rebuild and maintain situation awareness and cognitively engage with the situation. The RIS was developed by the DLR in collaboration with NLR and implemented in DLR's A320 Air Vehicle Simulator (AVES). The system emulates an A320 ECAM display by providing extra ECAM "risk pages" illustrated in Appendix B (adapted from Buch et al., 2017). The RIS aims to support the three principles in the following ways:

1. Time management: Risk overview pages (flight phases, aircraft characteristics) show the crew whether immediate action is required (which is not often) and that there is time to engage cognitively (which is more often the case).
2. Uncertainty management: Risk detail pages provide causal descriptions as well as action considerations that support crews in further diagnosing the situation and understanding the impact on the aircraft.
3. Contingency management: Risk overview pages allow flight crews to plan and identify risks in alignment with industry Threat & Error Management (TEM) concepts, developing more effective options and long-term plans.

Crews trained in the strategy and display solutions were provided similar strategy training, briefing and quick reference cards as the strategy-only crews, although these resources were all adjusted to include and explain the RIS as an integral part of the strategy execution. This included a modified quick reference card indicating which RIS pages may be most helpful in certain strategy phases (Appendix A).

3.2.3 Man4Gen validation study

The study hypothesized that the more training and support was provided to a crew in the form of strategy (training) and the RIS, the better their performance would be in managing safety-critical events. To investigate this, the study featured three test groups:

- A baseline group (BSL) which did not receive any training or RIS;
- A strategy only group (STG) which was trained in the new problem-solving strategy; and
- A strategy with display group (DIS) which received both strategy training and the RIS.

The groups were distributed across two experiment locations at NLR in Amsterdam and DLR in Braunschweig. Both locations featured BSL crews, the STG group was tested at the NLR, and the DIS group was tested at the DLR as they had the RIS integrated. At this point it must be mentioned that the study suffered from pilot recruitment, data collection and integration complexities which reduced the effective sample size considerably (discussed in Section 3.6). A Kruskal-Wallis paired ANOVA test indicated no clear differentiation between the above three test groups in terms of safety performance, likely attributable to the (very) small group sample sizes. However, reviewing performance data across the entire population (N = 14) indicated that there was considerable *within-group* variation [5]. For the current study, this resulted in a segmentation of the initial research hypothesis into two underlying causal pathways which potentially would allow for a stronger study design. The two resulting hypotheses are:

1. Crews trained in the strategy—with and without the supplementary RIS— (i.e., STG and DIS groups) will feature behaviors more closely aligned with the strategy, as compared to crews without training or RIS (i.e., BSL group). This will be evaluated by analyzing *between-group* differences in behavioral data.
2. Crews exhibiting behavior more closely aligned to the strategy will feature higher performance scores, and vice versa. As all crews (regardless of their grouping) provide both behavior and performance data, this can be investigated through correlation analysis considering all crews *as a single population*.

Both analyses will be done at two levels of abstraction: the strategy phase level (with six phase levels), and the strategy principles level (with three principle levels). By investigating the strategy's effectiveness at both phase- and principle-levels, this experiment may be able to discriminate whether the strategy's phases provide useful guidance to flight crew or—in the event they are not clearly effective—whether the underlying principles remain valid.

3.3 Method

3.3.1 Setup and scenario

The experiment featured 14 crews in total [6], each crew consisted of two pilots type-rated for either the Airbus A330 (NLR setup) or A320 (DLR setup). Three crews were not provided any Man4Gen training (BSL group), seven crews were trained in the Man4Gen strategy only (STG group), and four crews were trained in both the strategy as well as being

[5] Coefficient of quartile variation >30%

[6] Originally there were 17 crews, but three crews (all in the BSL group) did not feature complete behavioral datasets to be included in this evaluation. This reduced the BSL group from six crews to three crews.

provided the RIS (DIS group). As a general remark, the small number of crews warrants careful exploration of the study's results, both statistically and qualitatively.

All crews were provided a simulated flight scenario either at NLR or DLR, which was designed to provide similar challenges as two other scenarios explored in previous Man4Gen studies (Rankin et al., 2016). The scenario provided crews with a complex, ambiguous situation which challenged their ability make sense of the situation and regain cognitive control. This chapter will not detail the method used to design this scenario; it only provides a synopsis of the scenario. Readers are referred to (Niedermaier et al., 2018) for a detailed account of the scenario design method. Below is a synopsis of the scenario, a visual schematic of the scenario is provided in Appendix C.

> *"The scenario entails an Airbus A320 or A330 (depending on the simulator), and begins with a take-off, climb out and departure with bad weather approaching in the vicinity, with the redundant Flight Control Data Concentrator (FCDC2) and Slat Flap Control Computer 2 (SFCC2) flight control computers faulted but reported as (legal) Master Minimum Equipment List (MMEL) conditions. These flight control computer faults have no effect on the scenario, other than to potentially misguide crews at the start of the scenario. During climb-out the Engine #2 (ENG2) oil temperature begins to rise. Upon reaching advisory temperature, the crew may elect to reduce thrust or (if referred to the Flight Crew Operating Manual—FCOM) disconnect the Generator 2 (GEN2) or Integrated Drive Generator (IIDG2) electrical systems, doing so will be rewarded with a lower/stable oil temperature. However, maintaining climb thrust will further increase oil temperature and trigger an ECAM warning requesting an ENG2 shutdown. 14.5 minutes into the flight the aircraft is struck by lightning which has two effects. One effect is a Thrust Lever Fault of Engine #1 (ENG1), which reduces ENG1 thrust control to binary control: either Max Continuous Thrust (MCT, ~96% N1) or, when either slats or gear are extended, idle thrust. Normally this is not a problem because a functioning autothrust system will be able to still control the engine at all power levels. However, this option is not readily available, as the lightning strike also affects the autothrust system whereby the autothrust "speed mode" does not function (and disengages autothrust), however autothrust "thrust mode" does work. Speed mode is activated in the following common autoflight modes (and a few others): vertical-speed, ALT, GS, FINAL. If any of these modes are selected by the crew, autothrust disengages and ENG1 control returns to binary manual control, controllable solely by slat and gear positions. Crews experience the critical lightning strike about two minutes before they reach a point equidistant from three diversion fields, including their departure*

(home) field which has become relatively unsafe due to severely deteriorating weather (approaching thunderstorms). They are subsequently left to decide about a return or diversion and manage the energy challenges for the descent and approach."

As this was a departure scenario, crews had enough fuel and time (several hours) to address the problem and were briefed as such to take as much time as they need. The scenario was developed to work for both the DLR A320 AVES simulator and the NLR A330-configured Generic Research Aircraft Cockpit Environment (GRACE) simulator. For this reason, the scenarios in these sims featured slightly different navigation and endurance settings such that the challenges would be comparable for both an A320 and A330 setup. Except for a few differences in temperature limits, the thrust control systems and failure modes are identical between these two aircraft types.

The scenario was designed to challenge the crews around themes that are relevant for fourth generation aircraft operations and technology and provide the context for which the strategy and RIS has been designed. A comprehensive explanation of the design of this specific scenario can be found in Niedermaier (2018), Table 3.2 provides a summary of the designed effects.

Table 3.2. Scenario characteristics for specific effects on the crew (adapted from Niedermaier, 2018, Table 3)

Crew effect	Scenario characteristic
Ambiguous decision-making (uncertainty management)	▪ Return to the departure airport to facilitate maintenance vs. risk of airport closure due to approaching thunderstorms ▪ Most suitable diversion airports causing longer flight time with unreliable right-hand engine and increased fuel consumption ▪ Restart/keep right-hand engine after left-hand engine thrust lever fault although violating SOPs on ECAM
Surprise (manage immediate threats)	▪ Thrust lever fault usually not part of standard airline training ▪ Autothrust failure causing sudden disengagement of autothrust
Enhanced monitoring (uncertainty and contingency management)	▪ Detection of high oil temperature ECAM advisory on right-hand engine
Complex long-term planning (contingency management)	▪ Consideration of reduced go-around capabilities due to Full Authority Digital Engine Control (FADEC) behavior after slat extension in case of thrust lever fault ▪ Consideration of strong crosswind at some of the alternate airports

	▪ Consideration of thunderstorm movement to departure airport
	▪ Energy management with one engine in IDLE and one in MCT during cruise and descent and both engines almost in IDLE during approach
Uncertainty about system status / confusion (uncertainty management)	▪ Reliability of right-hand engine due to increased oil temperature
	▪ Unreliable autothrust function due to autothrust failure
Deep system knowledge (uncertainty and contingency management)	▪ Knowledge of the FADEC behavior after ENG 1 thrust lever failure

Crews in the two treatment groups (STG and DIS) were trained in the strategy. Training consisted of a briefing document explaining the philosophy principles and the strategy sent per email prior to the session, a two-hour training session to explain the strategy, applying it in several case studies, and practicing it in the simulator with a simple approach scenario featuring an extreme fuel leak. In this practice scenario, the aircraft remained well within range of its destination airfield, despite heavy fuel leakage, and provided crews with an opportunity to integrate the strategy into their operational routines. This allowed the crews from the two treatment groups to also use their quick reference card and familiarize with it in practice. The training and quick reference card provided to the DIS group was slightly augmented compared to the STG group, to highlight and reinforce the combined use of the RIS and strategy.

BSL crews did not receive any training in problem resolution, to allow for natural behavior. In order to mitigate experimental effects, the BSL crews still flew the fuel leak approach practice scenario as an "acclimatization scenario" to ensure all crews in all groups had approximately the same amount of (familiarization) time in the simulator. All crews were also invited for a short free-flight session before the fuel leak scenario to become accustomed to the research simulators' peculiarities. Figure 3.1 illustrates the sequence of simulation scenarios in this study.

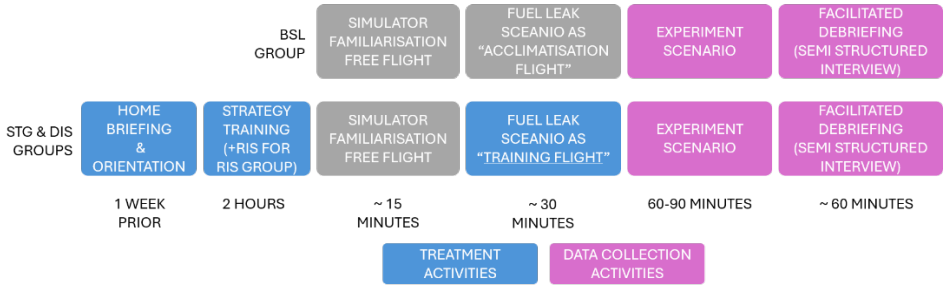


Figure 3.1. Illustration of the study setup

3.3.2 Performance and behavior dependent measures

Two main datasets were collected from the scenario runs: crew performance data and crew behavioral data. Crew performance data comprised of benchmarking crew actions against a list of actions that would maximize flight safety. This method will be explained in section 3.3.2.1. Crew behavioral data consisted of behavioral observations which have been parsed according to strategy phases. At a high level, the parsing consisted of:

1. organizing observations into strategy phase instances;
2. defining phase activity blocks based on these instances;
3. correcting phase activity for a multitasking index (resulting in duration measures); and
4. extracting attentional redistribution between block (resulting in sequence measures);

All data collection was done according to NLR and DLR experiment ethics policies, including a data protection and privacy policy form signed by both subjects and a research team member, and all data being stored responsibly and securely. Participants were allowed to revoke their data being collected or used and could also request their data in accordance with NLR and DLR privacy policies. Experimenters took great care to avoid any negative training, and to provide participating crew members with ample opportunity to discuss any negative impact the experiment may have had on their professional confidence (no negative impact was established for any participant).

3.3.2.1 Crew performance measures

The first dataset, crew performance, was defined as a crew's ability to maintain a high level of safety throughout the scenario. Crews were not explicitly instructed to maximize safety, as the study was semi-naturalistic and aimed to replicate authentic behaviors and prioritizations, except for the strategy and display treatments. Crew performance was measured using the Desired Flight Crew Performance (DFCP) method (Field et al., 2016). The DFCP method was developed earlier in the Man4Gen project and essentially provides a useable performance benchmark to compare crews in a scenario in which there are multiple solution pathways. The DFCP method consists of a list of very scenario-specific desired decisions and actions which are marked as observed (performed) or not. A perfect score is not possible as a crew cannot simultaneously engage in all different solution pathways. In this study the DFCP score was used as a relative measure, and not as an absolute indication of safety performance. The reader is directed to (Field, 2016) for a detailed explanation of the DFCP method.

All DFCP items (30 for this scenario, see Appendix D) were weighted equally with the final performance score being the sum of all correctly executed items. High performing crews would have high scores, and low performing crews would have lower scores. The DFCP items were scored by the main author post-hoc with the use of multi-source video recordings (including display recordings), supported by cross-examination of the crew decisions, thoughts and considerations in post-scenario interviews (also recorded). Performance scores were validated with an active-duty type-rating instructor and examiner to reduce observer bias. Next to the total DFCP score variable, four sub-DFCP variables were defined, based on specific challenges and complexity factors in the scenario. The reason for this is that this accounts for different constellations of local high and low performance which may otherwise cancel out in a total DFCP score (e.g., one crew is effective in challenge A and weak in challenge B, and another crew vice versa). Sub-performance indicators (sub-PIs) were divided according to four main complexity factors and challenges in the scenario:

- ENG2 DFCP: Engine 2 management (oil temperature rising) (11 DFCP items),
- ENG1 DFCP: Engine 1 management (thrust lever fault) (7 DFCP items),
- ROUTE DFCP: Route management (weather, destination choice, descent) (8 DFCP items), and
- COMM DFCP: Communication (with ATC, dispatch) (7 DFCP items).

Note that a few DFCP items overlap in challenges and contribute to two sub-PI's. Two remaining DFCP items could not be readily classified into the above sub-PIs but were still part of the total DFCP score.

3.3.2.1.1 Crew behavioral measures

The second dataset consists of crew behavioral measures, which were generated by an observation form that followed the structure of the strategy phases and associated activities. The observation form included a total of 23 behaviors, distributed across the six phases. These were supplemented with six additional indicators whether a phase was verbalized (i.e., did the crew explicitly mention the use of any part of a phase such as "*let's first make a short-term plan*" or "*we've done the first phase, the aircraft is stable*"). The verbalization indicators were used in the Man4Gen study to evaluate the training and investigate how often strategy-trained crews retrieved and used the quick reference card. For this study, those verbalization indicators are not used as performance indicators, as the baseline crews would then suffer an unfair disadvantage because they were neither trained in the strategy, nor were provided the quick reference card to verbalize. The other 23 behavioral indicators were generic enough, however, to observe in all crews regardless of their training. All behaviors were marked along a timeline for all crews. Strategy behavior markers were recorded and timestamped by the main author post-hoc with the use of the same video recordings as used

for the performance data using Noldus© software, generating data as illustrated in Figure 3.2. Behavioral analysis was validated by a secondary observation performed by a human factors specialist. Note that most behavior marking occurs after about 14-15 minutes of runtime. This is because scenario non-normal operations, malfunctions and complexities only start developing around this time.

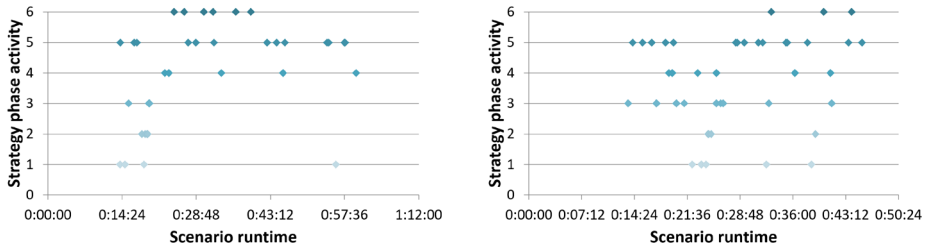


Figure 3.2. Examples of behavioral mapping of the strategy phases over time, for two different crews. The pattern on the left is of a crew with a high DFCP score, the pattern on the right of a crew with a low DFCP score.

Behavioral patterns are investigated at two abstraction levels, pertaining to the six-phase strategy as well as the three underlying principles. For the principles analysis, the behavioral indicators among the six phases are clustered as follows: Behavioral observations in strategy Phases 1, 2 and 3 are all considered TM, Phase 4 is UM and Phases 5 and 6 are considered CM. The purpose of this a-priori composite variable aggregation is to reduce noise that may stem from nuances in applying the strategy phases (e.g., P2 "manage immediate threats" may or may not occur if the crew does not have or see an *immediate* threat, but general stabilization/TM may well occur). Composing a smaller set of aggregate indicators from the raw observations also reduces Type II errors that may be present in the large number of measures at the strategy phase level.

After mapping and classifying the basic instances of observed behaviors, the data were processed to distil two subsets of behavioral data:

- phase/principle *duration* (how much time was spent on a particular phase/principle), and
- phase/principle *sequence* (in which order were phases/principles demonstrated).

The methods employed to process the raw behavioral data into these two characteristics are explained in the next two subsections.

3.3.2.1.2 Behavior duration analysis method

For both duration and sequence analyses, the instance-based behavior data (as seen in Figure 3.2) had to be converted into segments of activity for each of the phases. Adding

these segments provides a heuristic approximation of the total time (minutes) that a crew spent on each phase (which will subsequently be corrected for multitasking). As the exact start and end timing of activity in a specific phase was (obviously) not declared by pilots, nor is it easily identifiable from videos or debriefings, it was chosen to define phase activity time blocks using a mathematical heuristic method dubbed the "blocking algorithm".

The blocking algorithm defines phase activity segment start- and end-times by working chronologically along a given phase, using a time criterion to group observations. The time criterion for this experiment was set at three minutes, as will be explained below. This process begins at the first instance of observing a phase-related behavior, and from that timestamp on determines if the next observed behavior of that phase occurs within the subsequent three minutes. If it does, the block is extended to this next point, and the process checks again if there is a subsequent third observation of the same phase within three minutes of this second point, and so on. If no observation of this phase's activity occurs within the next three minutes, the timestamp of the last observation point *plus three minutes* is set as the end time for this block. Figure 3.3 illustrates the blocking algorithm.

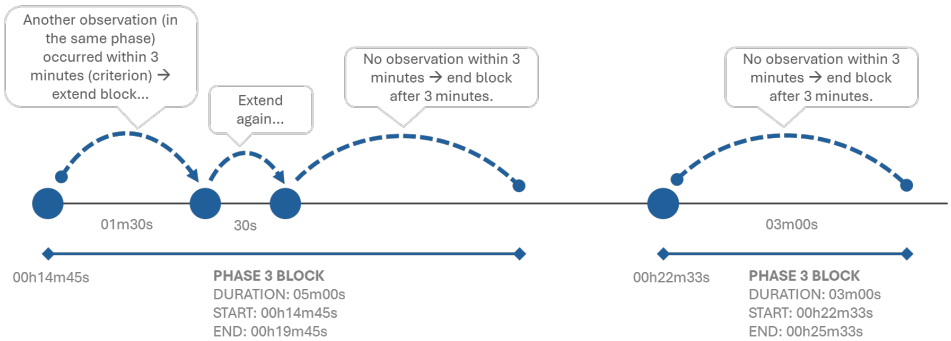


Figure 3.3. Blocking-algorithm to cluster phase behavior observation instances (3-minute time criterion)

The time criterion is determined using a multitasking index. This is calculated by summing all blocks across all phases (total "modelled" behavior) and dividing in by each sample's total scenario runtime, providing a multitasking measure for each sample. A shorter time criterion will result in smaller, erratic time blocking which reduces phase concurrence (multitasking), but may not account for continued cognitive engagement in a phase despite sparse observations. A longer time criterion will result in larger, more continuous time blocking which may increase phase concurrence beyond what may be assumed as cognitively reasonable for a two-person crew. These metrics are calculated from the behavioral data and are unique to the dataset collected for these crews in this scenario. Different time criteria were tested as shown in Table 3.3.

Table 3.3. Comparison of several time criteria for the blocking-algorithm

Time Criterion	1.5 MIN	2 MIN	2.5 MIN	3 MIN	4 MIN	5 MIN
Average multitasking	0.92	1.14	1.33	1.50	1.80	2.06
Highest multitasking	1.25	1.53	1.75	1.97	2.34	2.70
Lowest multitasking	0.61	0.75	0.87	0.99	1.17	1.33
σ multitasking	0.197	0.223	0.245	0.268	0.318	0.369

Based on the assumption that each crew member can only actively be engaged in one task rather than truly multitasking multiple tasks concurrently (Loukopoulos et al., 2009), the three-minute criterion is selected as the average multitasking index lies between one and two. Smaller criteria show more gaps in the blocking models (multitasking indices less than one), while longer criteria begin to exceed a multitasking limit of two. The multitasking data presented below are calculated across the entire population (N=14). A Kruskal-Wallis between group analysis of multitasking indices indicated no significant differences between the BSL, STG and DIS groups in terms of their multitasking index, indicating that this three-minute blocking model may be used similarly for the entire study population.

This blocking model provides duration measures for each of the six phases (the runtime sum of all blocks per phase). This is the raw (observed) duration in each phase. These raw measures are further transformed into two other sets of phase duration measures. The first is correcting the duration for multitasking to account for finite cognitive capacity, by dividing the duration of overlap across the number of phases concurrently active, as illustrated in Figure 3.4.

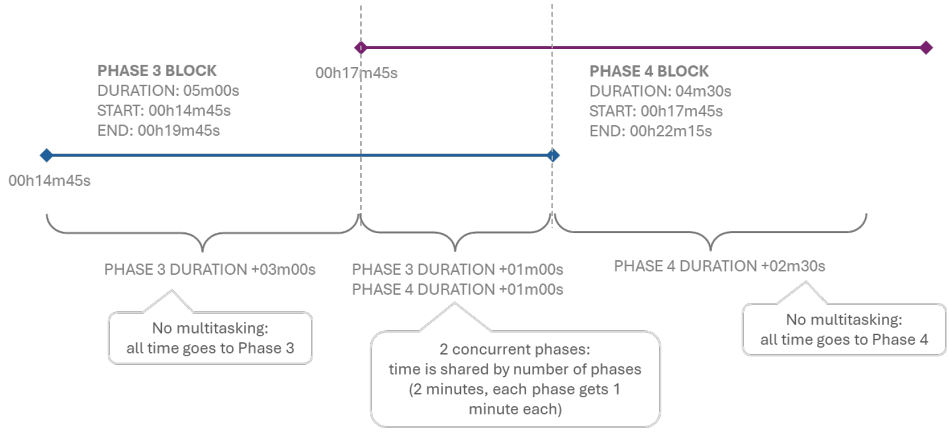


Figure 3.4. Example of phase duration multitasking correction

This avoids an over-valuation of time "active" in a phase if that phase is constantly run parallel to other activities, thus providing a proxy for a crew's actual division of their attention across the phases. The other correction is subsequent expressing these multitasking-corrected durations as a percentage ratio. This provides a proxy for crew phase preference and selection, as well as allows for a comparison of crew duration patterns despite different runtimes. The six observed, six corrected and six ratio duration measures provide different *lenses* to observe relations of behavioral measures to each other and to performance scores. By observing how correlations may disappear (e.g., when correcting for multitasking) or become stronger (e.g., when correcting for time with ratios), the results can segregate between particularly strong and weak corrections [7]. It is expected that (relatively) more attention spent on later phases (Phase 4, 5 and 6) and (relatively) less on earlier phases (notably Phases 1 and 2) will correlate positively with increased DFCP performance across all DFCP sub-PI's. Figure 3.5 illustrates the transformation of these correlation matrices. Note that the study will not correlate different transformations of the same data (e.g., correlating observed data with ratios).

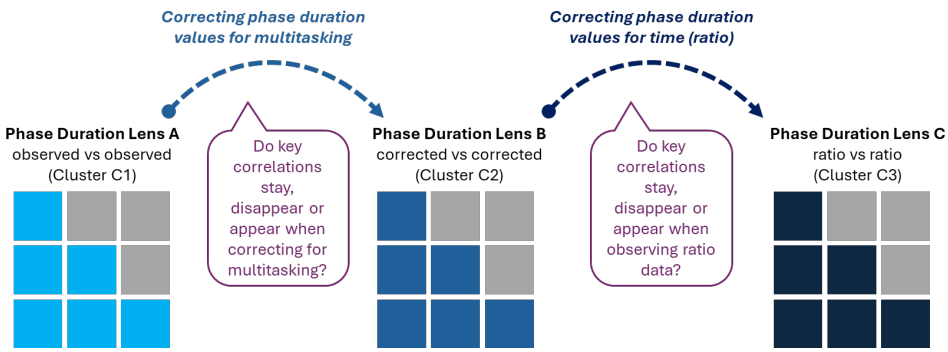


Figure 3.5. Illustration of the two transformations of the correlation matrices to Lens B and Lens C, respectively.

At a higher level of abstraction, the three philosophy principles were also characterized using the same time criterion of three minutes, resulting in three duration measures (one for each principle), which are also subject to the multitasking and ratio transformations. It is expected that (relatively) more attention spent on UM and CM, and (relatively) less on TM will relate to higher DFCP performance.

[7] As a general observation: the transformation to *ratio* will expect to dampen positive correlations as the data is now constrained to 100%. Conversely, negative correlations may become stronger. That is the value of such a relational lens.

3.3.2.1.3 Behavior sequence analysis method

Next to phase/principle duration, the study also investigated phase/principle sequencing. The sequence analysis characterization uses the same data structured by the blocking algorithm, and tracks when crews switch their activity from one strategy phase to another, and to which phase they switch. This provides an indication whether crews are following one strategy phase with the subsequent strategy phase, or whether they are skipping phase steps, or going back to previous phases. These indications can then be compared to DFCP scores to identify sequence patterns that increase or decrease safety performance. The focus lies on the transition vector between phases, to be able to model the order of a crew's activity. This method assumes that cognitive load and attention is divided equally between all concurrently active phases and therefore can indicate how much attention is being shifted from one phase to another when a given phase time block activates or terminates. The equal distribution assumption is based on the actual co-occurrence of activity in multiple phases and relatively short time criteria (three minutes) to end (or maintain) phase activity. Characteristic behavioral patterns in attention switching between phases can then be calculated along the scenario runtime.

Phase switching is evaluated numerically by modelling the amount of crew attention transferred from one phase to another. Figure 3.6 illustrates this concept for a given constellation of phase time blocks. Observe in Figure 3.6 the small arrow on the left which indicates that when the Phase 3 time block activates (with Phase 2 still active), the sequence algorithm assumes that 50% of the attention is now shifted from Phase 2 to Phase 3, numerically indicated as J_{23}=50%. The diagram at the bottom of Figure 3.6 shows how the attention is distributed between different activation and termination events. Similarly, when Phase 5 is activated while Phases 2 and 3 remain active, attention must be redivided from two phases to three phases, resulting in two equal attention switches of about 17% from Phases 2 and 3 to Phase 5.

Figure 3.6 also illustrates how the sequence algorithm redistributes attention when a phase block terminates. When Phase 5 terminates, this results in a redistribution of attention to the three active phase (2,3 and 4), which increase from 25% to about 33% attention demand per phase. Similarly, when the Phase 2 block terminates, 50% of attention is moved from Phase 2 to Phase 4, which then has 100% of the crew's attention. When no phases are active, attention is set to 0% (the next phase to activate receives 100% attention). However, the last known active phase is still used as the algorithmic departure point if a different phase activates after a period of zero phase activity. It is recognized that the phase switching method assumes that multiple tasks may be concurrently active, at times more than the maximum assumed multitasking index of two for a crew of two persons. However, given

that the average multitasking index was found to be 1.5 for this choice of time criterion (see Table 3.3), such instances will be limited and of relatively short duration.

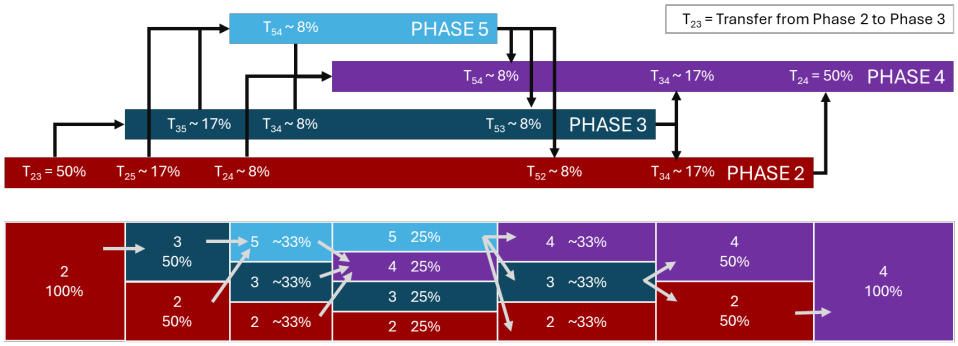


Figure 3.6. Phase switching example – illustration of attention redistribution (%) between phases

Results of the sequencing algorithm are then organized in a matrix that tracks the runtime sum of all 30 possible attention transfers between phases. The final variable value for each transfer type is the cumulative runtime sum of attention switched as such. Table 3.4 illustrates how the example block sequence shown in Figure 3.6 is translated to cumulative sequence data. As Figure 3.6 only depicts a basic, limited set, a complete scenario may feature cumulative attention switching values exceeding 100% for one or multiple switches.

Table 3.4. Example of cumulative phase switching based in the illustration in Figure 3.6.

Figure 3.6 Example	**From P1 (TM)**	**From P2 (TM)**	**From P3 (TM)**	**From P4 (UM)**	**From P5 (CM)**	**From P6 (CM)**
To P1 (TM)						
To P2 (TM)			17%		8%	
To P3 (TM)		50%			8%	
To P4 (UM)		58%	23%		16%	
To P5 (CM)		17%	17%			
To P6 (CM)						

The 30 phase transfers were further grouped into five characteristic strategy phase sequencing patterns defined with respect to the Man4Gen strategy phase utility, and three characteristic principles sequencing patterns for strategy principles utility. These eight characteristics are illustrated in Figure 3.7 and described in more detail below.

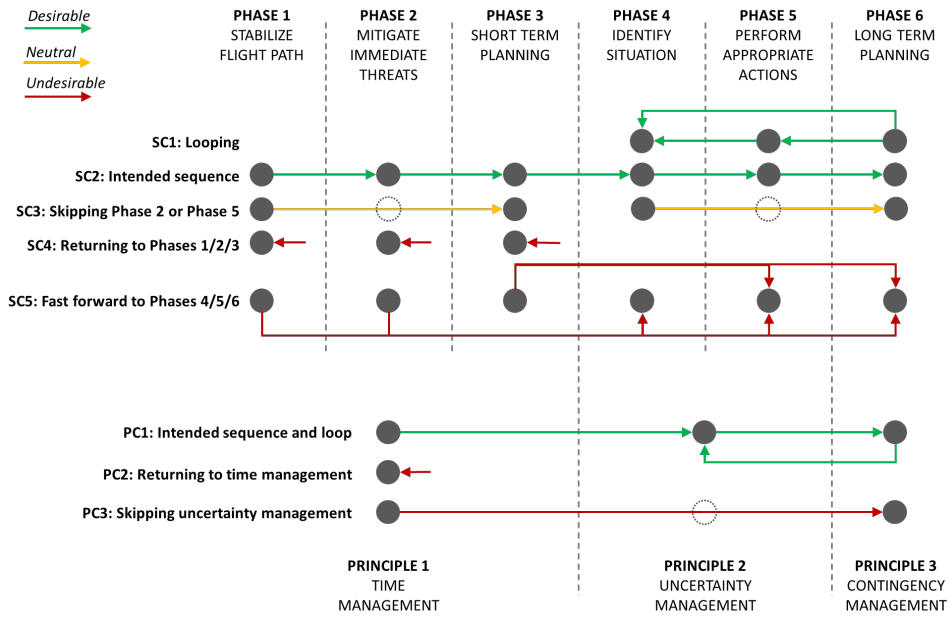


Figure 3.7. Overview of strategy and principle sequence characterizations

Strategy Phase Characteristic 1 (coded as SC1) refers to the crew repeatedly engaging in cyclic sensemaking with contingency management, by cycling through Phases 4, 5 and 6. While this is chronologically not the first behavior to observe, it is hypothesized to have the strongest effect on performance. SC2 refer to following the sequence of phases as intended, and monitors how often a crew moves from one phase to the subsequent phase of the strategy. It is expected that these two characteristics (SC1 and SC2) are positively correlated to higher DFCP performance scores. This is based on the expectation that more time spent in later phases (i.e., Phases 4, 5 and 6) will improve crew sensemaking of ambiguous situations and identify better/safer contingency plans. Note that the intended sequence is not to cycle back to Phase 1 as that would indicate ineffective short-term time management. Rather, the intention is that after reaching Phase 5 or 6, the crew returns to Phase 4 for continued monitoring and management of uncertainties and risks (SC1).

SC3 measures whether a crew skips short-term actions (Phase 3) or appropriate recovery actions (Phase 5). By skipping such phases, crews forfeit an opportunity to learn and engage with the system. On the other hand, sometimes there may not be any immediate or recovery actions applicable. This behavior is expected to correlate slightly negatively with DFCP performance scores. SC4 and SC5 refer to skipping the initial stabilization phases or returning to them from later phases, respectively. As the Man4Gen strategy was designed to creating stability and time at the start of a developing situation to provide problem-solving

space, it is expected that the behavior of skipping either these initial phases or a necessity to return to them later would correlate negatively with DFCP performance scores

Characterization Matrix (example)	From P1 (TM)	From P2 (TM)	From P3 (TM)	From P4 (UM)	From P5 (CM)	From P6 (CM)
To P1 (TM)		4	4	4	4	4
To P2 (TM)	2		4	4	4	4
To P3 (TM)	3	2		4	4	4
To P4 (UM)	5	5	2		1	1
To P5 (CM)	5	5	5	2		1
To P6 (CM)	5	5	5	3	2	

SC 1 = Looping between Phases 4,5 or 6 (SCSC1) - Desirable

SC 2 = Intended sequence from P1 to P6 (SCSC2) - Desirable

SC 3 = Skipping Phase 2 or Phase 5 (SCSC3) – Neutral

SC 4 = Going back to Phases 1,2 or 3 (SCSC4) – Not Desirable

SC 5 = Fast forwarding to Phases 4, 5 or 6 (SCSC5) – Not Desirable

Figure 3.8. Strategy phase-switching matrix showing the five characteristic behavioral patterns

Each characteristic is numerically expressed (for each crew) as the total scenario runtime sum of all transfers allocated to that SC group (as illustrated by Figure 3.8). Each SC sum is normalized by dividing that sum by the number of transfers in each SC group to account for equal distribution likelihoods. The total runtime sum for each SC is expected to provide sufficient variation to validate whether desired and undesired characteristic sequences have positive and negative effects on DFCP performance, respectively. These normalized runtime sums are also transformed to an additional set of characteristic ratios as a proxy to crew preference and selection of certain sequencing strategies and allow comparison of behavioral distributions despite different runtimes.

The total scenario runtime sum of a characteristic sequence may feature different underlying patterns. Equivalent total runtime sums may consist of many smaller transfers (i.e., frequent task switching) or fewer but larger transfers (i.e., more task focus and less switching). Figure 3.9 illustrates four possible crew modalities in characterization sequencing.

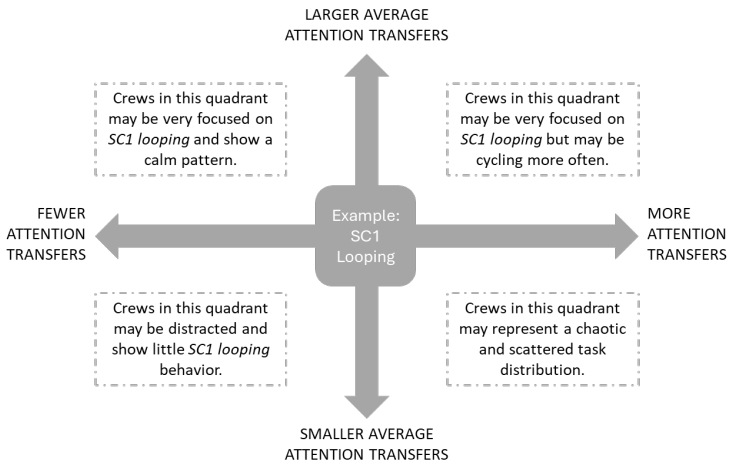


Figure 3.9. Description of four characterization modalities

However, given the limited number of samples and highly granular nature of such sub-characterizations (i.e., number of attention transfers and average size of attention transfers), this study will only broadly and qualitatively explore whether crews show large differences in modalities and if sequencing measures should account for this. Subsection 3.4.2 will provide such a characterization modality plot for each SC measure for exploratory discussion. For the purpose of hypotheses, correlations and regression analyses, only the total sum and respective ratio measures will be used. This is also to avoid possible strong multicollinearity between the size and number of transfers (i.e., smaller sized transfers will often coincide with more frequent transfers as a result of the phase switching algorithm).

Table 3.5 illustrates a basic example of these variable transformations and normalization from the blocking described in Figure 3.6 and Table 3.4. Figure 3.10 shows an example of different strategy phase characterizations over the runtime of a scenario.

Table 3.5. SC measures extracted from the example in Figure 3.6

SC Measure	**Normalized cumulative transfer**	**Normalized total transfer count**	**Average Transfer size** [8]
SC1	16% ÷ 3 = 5.3%	2 ÷ 3 = 0.67	16% ÷ 2 = 8%
SC2	73% ÷ 5 = 14.6%	3 ÷ 5 = 0.6	73% ÷3 = 24.3%
SC3	0% ÷ 2 = 0%	0 ÷ 2 = 0	0% ÷ 0 = 0%
SC4	33% ÷ 12 = 2.75%	3 ÷ 12 = 0.25	33% ÷ 3 = 11%

[8] Clarification: The average transfer size is not normalized. Dividing the normalized total runtime sum by the normalized total transfer counts provides the average transfer size (normalization factors cancel out).

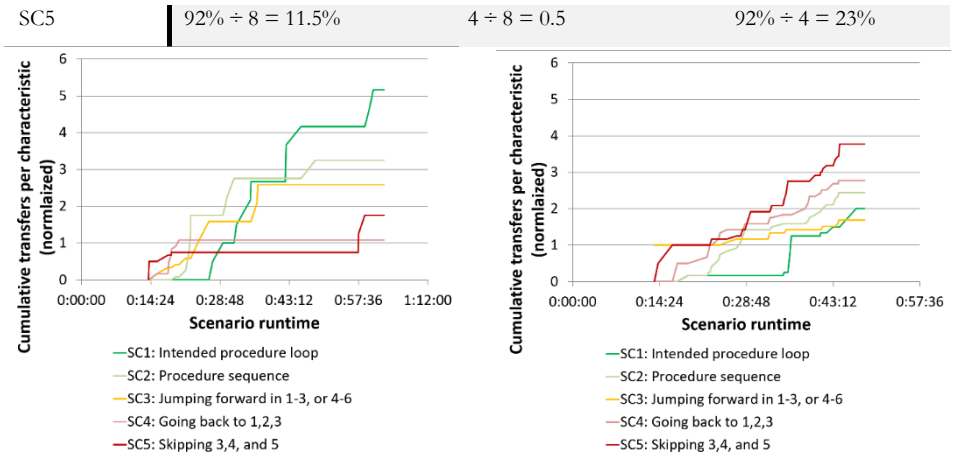


Figure 3.10. Examples of output of the characterization analysis, clearly illustrating different behavioral patterns. The pattern on the left is of a crew with a high DFCP score, the pattern on the right of a crew with a low DFCP score.

At a higher abstraction level, crew behaviors were also characterized into the three philosophy principle characterizations. This was performed using a philosophy principles characterization algorithm and featured a smaller characterization matrix with only six transfers, as illustrated in

Figure 3.11. Three behavioral characterizations are modelled: Principle Characteristic 1 (PC1) describes the desired sequence and looping of principles (similar to a combination of SC1 and SC2) and was expected to correlate positively with DFCP performance scores. PC2 and PC3 are undesired characteristics, respectively describing a return to TM or skipping UM, both of which were expected to correlate negatively with DFCP performance scores. All PC measures are also transformed to relative ratios to indicate crew preference for sequence patterns.

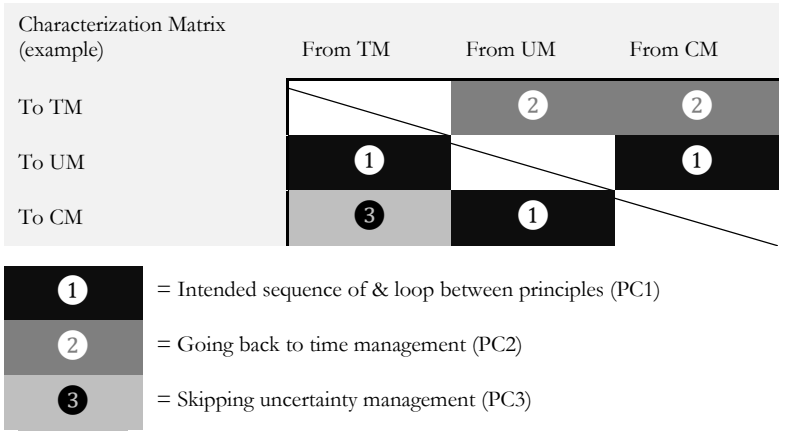


Figure 3.11. Principles-switching characterization matrix

At both levels of abstraction, the phase switch characterization approach is limited to analyzing the amount of switching between any two phases. While it may be possible to also model and track all 64 unique states (i.e., all possible combinations of phases co-activation) that a crew can be in, the derivative nature of such a behavioral model, the increasing number of assumptions combined with the low number of samples (N=14) would quickly render any analysis invalid.

3.3.2.2 Summary of measures

The study will feature five performance measures: their total DFCP score as well as the four sub-PI's for Engine 2 management (ENG2 DFCP), Engine 1 management (ENG1 DFCP), route management (ROUTE DFCP) and communication (COMM DFCP). Each (sub-)PI score is expressed as percentage of the total achievable DFCP (sub-) PI. The duration and sequence measures at both the strategy and principle levels result in ten distinct *clusters* of behavioral measures. Each cluster can be regarded as a behavioral *lens* through which interactions may be studied. Table 3.6 provides an overview of these clusters, their rationale and their ID (to facilitate results reporting).

Table 3.6. Overview of and rationale for behavioral measures clusters

ID	Cluster Name	Variables	Rationale
C1	Strategy Phase Duration - observed	6 measures *P1-P6 Duration (observed)*	Ability to contrast with multi-tasking corrected durations to observe phase concurrence
C2	Strategy Phase Duration - multitasking corrected	6 measures *P1-P6 Duration (corrected)*	Measures that account for the reality of finite cognitive capacity
C3	Strategy Phase Duration - multitasking corrected (% ratio)	6 measures *P1-P6 Duration (%corrected)*	Ability to contrast patterns in durations and account for runtimes (e.g., crew preferences)
C4	Strategy Characterization – normalized	5 measures *SC1-SC5*	Fair comparison of key sequence patterns (normalized to category number of possible transitions)
C5	Strategy Characterization – normalized (% Ratio)	5 measures *SC1-SC5 (%)*	Ability to contrast patterns in durations and account for runtimes (e.g., crew preferences for sequences)
C6	Principle Phase Duration - observed	3 measures *TM / UM / CM duration (observed)*	Same as C1, but for principle level
C7	Principle Phase Duration – multitasking corrected	3 measures *TM / UM / CM duration (corrected)*	Same as C2, but for principle level
C8	Principle Phase Duration – multitasking corrected (% Ratio)	3 measures *TM / UM / CM duration (%corrected)*	Same as C3, but for principle level
C9	Principle Characterization - normalized	3 measures *PC1-PC3*	Same as C4, but for principle level
C10	Principle Characterization - normalized (% ratio)	3 measures *PC1-PC3 (%)*	Same as C5, but for principle level

3.3.3 Study hypotheses

Two main hypotheses will be investigated:

Hypothesis H_1: Between groups analysis: Crews trained in the strategy both with and without the RIS (STG and DIS groups) demonstrate behavior more aligned with the Man4Gen strategy, than control group crews (BSL group). This means (relatively) more attention spent on Phases 4-6 (Principles: UM and CM) than Phases 1-3 (Principles: TM), a (relative) increase in SC1 and SC2 patterns (Principles: PC1) and a (relative) decrease in SC3, SC4 and SC5 patterns (Principles PC2 and PC3).

Hypothesis H_2: Within-group analysis: Crews demonstrating (relatively) more behavior in line with the Man4Gen strategy (as described in H_1), will feature higher DFCP/safety performance in the test scenario.

3.3.4 Data analysis methods

The study will analyze data in four ways:

Analysis I. Relations between performance measures to confirm sub-DFCP validity (stronger measures to evaluate H_2);

Analysis II. Relations with and between duration and sequence measures at the strategy phase- and principle-level, respectively (stronger measures to evaluate both H_1 and H_2);

Analysis III. Effectiveness of training in changing STG and DIS group behaviors at both the strategy phase- and principle-level (testing H_1);

Analysis IV. Relations between behavior and performance measures at the strategy phase- and principle-level (testing H_2);

Analysis I investigates DFCP measures for intra-DFCP correlations and sub-DFCP predictor validity (i.e., can the sub-measures adequately proxy for the total score). This is done to ensure the set of performance markers are mutually independent while representing the full scope of performance. Bivariate correlations (Kendall Tau) will identify whether the sub-DFCP measures relate to Total DFCP and/or to one another. A linear regression analysis will be performed to validate the sub-DFCP's as valid constituents for Total DFCP performance. The limitations of linear regression modelling and the risk of overfitting and Type I errors are appreciated. As the initial assumption is that all sub-DFCP variables have predictive power, a backwards stepwise regression is considered the preferred approach, reducing overfitting by removing those measures that contribute the least to the model's prediction power. The measure removal criterion is set at p=0.1 [9], monitored for a low Variable Inflation Factor (VIF) (less than 10 indicates low multicollinearity) and highest Adjusted R^2 to prevent underfitting and possible Type II errors (i.e., removal of measures with predictive value).

Analysis II features a comprehensive review to identify and validate correlations between various behavioral measures. This is done to reinforce the behavioral measures model before Analysis III and IV test the two study hypotheses. Duration and sequence behavioral measure sets are initially examined independently (i.e., duration measures

[9] Twice the minimum significant level of 0.05 is customary (Hosmer & Lemeshow, 2000)

correlated with other duration measures), across the different lenses. Robust correlations are those that appear in multiple correlation lenses (e.g., observed, corrected and ratio measures). Subsequently, correlations between duration and sequence clusters are identified and evaluated for their robustness across the data transformations. Awareness of strong relations between behavioral measures and possible covariation between measures will provide additional contrast when evaluating training effectiveness and behavior-performance correlations.

Analysis III tests H_1 by evaluating differences between the three groups, across all ten behavioral measures clusters. This is done by a Kruskal-Wallis (KW) non-parametric test (with a Benjamini-Hochberg (BH) correction, explained below) to identify significant between-group differences. Significant differences will feature post hoc pairwise testing using Dunn's test (also BH corrected) to identify which groups differed significantly.

Analysis IV tests H_2 by investigating the effectiveness of the Man4Gen strategy by investigating the correlation between behaviors and performance across all crews as a single population. Analyzing all crews as a single group for H_2 is appropriate as the variation of crew behavior and performance (regardless of whether this is random or due to training) provides better data stratification for correlation analysis, as well as a larger sample size. Each performance (sub-)PI will be compared to each behavioral variable by means of a linear bivariate correlation analysis using Kendall Tau to account for non-parametric datasets. BH correction is applied to each analysis cluster (e.g., DFCP vs.C1…C5…C10 etc.).

As a general caution given the low sample size, normality is not assumed given the low power of Shapiro-Wilk tests for N=14. Non-parametric ranking tests will be used to reduce sensitivity to outliers and possibly unequal variances, particularly in small group sizes (e.g., N=3). Furthermore, given the large number of behavioral measures, *p*-adjustment will be consistently applied to account for Type 1 errors (false positives). Adjustment will occur at the cluster level as different cluster of measures represent distinct relations with performance or other behavioral measures. As single cluster-level analysis can feature up to 30 correlations (e.g., 6 duration x 5 SC measures), a Benjamini-Hochberg (BH) False Discovery Rate (FDR) correction is applied to provide a progressive *p*-adjustment that is more appropriate for explorative studies with multiple independent variables [10]. Significance is tested at the p=0.05 level, BH-corrected *p*-values are reported as *q*-values and also tested at the q=0.05 level, with the following effects strengths being reported:

[10] Where the alternative Bonferroni correction aims to reduce the family-wide error rate, BH aims to reduce the false discovery rate, which is generally more relevant and applicable for explorative research.

- $q \leq 0.05$: significant effect (***);
- $p \leq 0.05$ but $0.05 < q \leq 0.1$: near-significant effect (**);
- $p \leq 0.05$ but $q > 0.1$: trend (*); and
- $0.05 < p \leq 0.1$ but with large effect size or τ is large [11]: strong effect (Δ).

3.4 Results

This section presents the performance and behavioral measures as collected and their analysis. Subsection 3.4.1 inspects the data collected in terms of normality and validates whether the DFCP Sub-PI's are valid segmentations of the total DCP performance (Analysis I). Subsections 3.4.2 and 3.4.3 present behavioral data for the strategy phase and principle level respectively, where Subsection 3.4.4 presents the relations between behavioral measure through multiple correlation lenses (Analysis II). Subsection 3.4.5 presents the evaluation of training effectiveness (Analysis III, for Hypothesis H_1) and Subsection 3.4.6 will present the relationships between behavioral and performance measures (Analysis IV, for Hypothesis H_2). The original behavioral and performance data for all crews can be found in Appendix E and Appendix F.

3.4.1 Performance measures results and Analysis I

Figure 3.12 provides an overview of the DFCP scores. In general, crews have more difficulty (lower DFCP scores) in managing Engine 2 than Engine 1 and showed a large variation in Route Management scores. The Total DFCP scores (leftmost in Figure 3.12) showed less variation than the other four Sub-PI's. This may be explained as different scenario resolution strategies may vary in their prioritization of various sub-performance elements (e.g., engine versus route optimization), while still achieving a similar Total DFCP score. This was in part the reason to recognize sub-PI's.

[11] The following is considered as large effect sizes for tests used: Kruskal-Wallis 0.14; Dunn post-hoc test > 0.5; Kendall $\tau > 0.5$ ($\tau > 0.3$ is considered moderate).

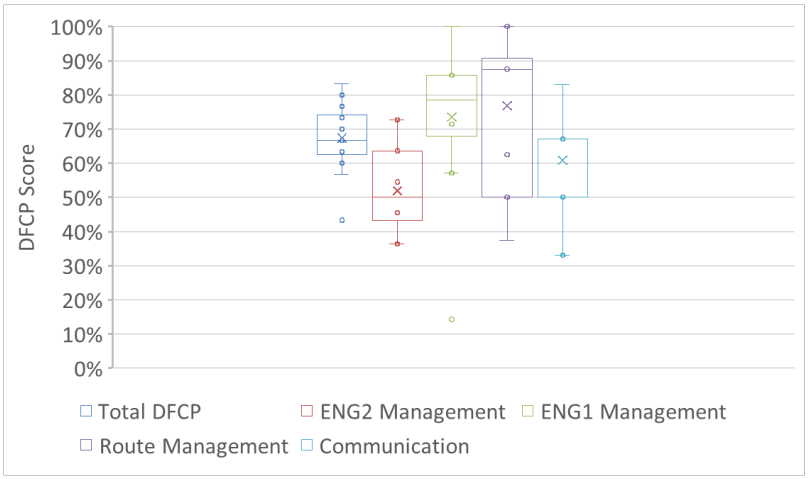


Figure 3.12. DFCP performance measure box plots

Despite its weak power for low samples sizes, Shapiro-Wilks testing for DFCP data normality indicated that ENG1 (p=0.005), ROUTE (p=0.01) and COMM (p=0.002) data are not normally distributed, confirming the use of non-parametric testing for all performance-related comparisons. Kendall Tau non-parametric correlation test statistics in Table 3.7 indicate that the performance metrics ENG2, ENG1 and ROUTE sub-PIs are each correlated with the TOTAL DFCP score while also featuring no significant correlations between themselves.

Table 3.7. DFCP variable correlation analysis - Kendall Tau with p and (q)

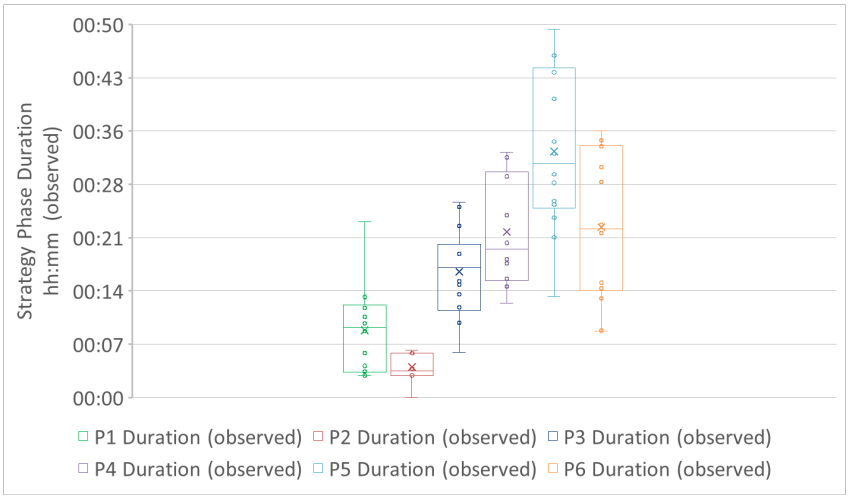
	TOTAL DFCP	**ENG2 DFCP**	**ENG1 DFCP**	**ROUTE DFCP**
ENG2 DFCP	0.597*** 0.006 (0.03)			
ENG1 DFCP	0.473* 0.034 (0.113)	0.288 0.211		
ROUTE DFCP	0.665*** 0.003 (0.03)	0.327 0.154	0.187 0.425	
COMM DFCP	-0.120 0.601	-0.0626 0.791	-0.314 0.195	-0.0657 0.786

*** Corrected p (q) is significant (q ≤ 0.05) (two-tailed)
* Uncorrected p is significant, trend (p ≤ 0.05; q > 0.1) (two-tailed)

A further backward stepwise regression [12] analysis (exit criterion p=0.1) indicated that, together, these same sub-PI's can indeed account for the majority of the TOTAL DFCP score variance (adjusted R^2 = 0.951), while featuring limited multicollinearity with all independent variables' VIFs close to one.

3.4.2 Behavioral measures results – Strategy phase level

The distributions of strategy phase duration measures (observed, corrected, and ratio) are shown in Figure 3.13. Overall, crews spent more time in later phases than in the initial phases. This is in general alignment with the scenario challenges which were more focused on uncertainty and contingency management, rather than immediate time management. It is notable that all crews spent very little time in Phase 2 ("Manage Immediate Threats"), and that there is considerable variation in the time spent in later phases, notably Phases 5 ("Perform Actions") and Phase 6 ("Long-Term Planning"). Correction of multitasking reduces the total durations (which is expected), but the general pattern appears consistent, also for the transformation to ratios.



[12] Ordinary Least Square (OLS) linear regression is permitted for non-normal variables as inspection of regression residuals shows a normal residual distribution, which is the case for this particular regression.

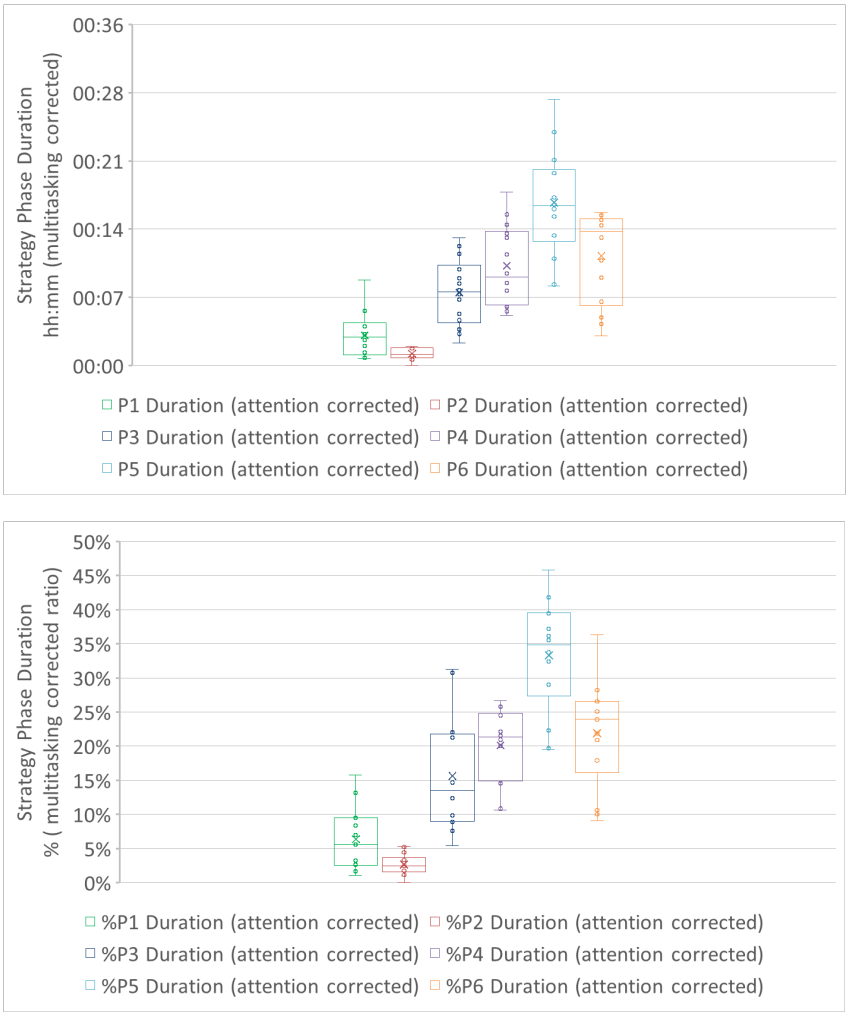


Figure 3.13. Strategy phase duration measures box plots (observed, corrected & ratio)

The distributions of strategy characterization measures are shown in Figure 3.14 (normalized runtime cumulative transfers, and ratio variant) and indicate that crews feature considerably more SC1 (desired) and considerably less SC4 and SC5 (undesired) behavior sequences.

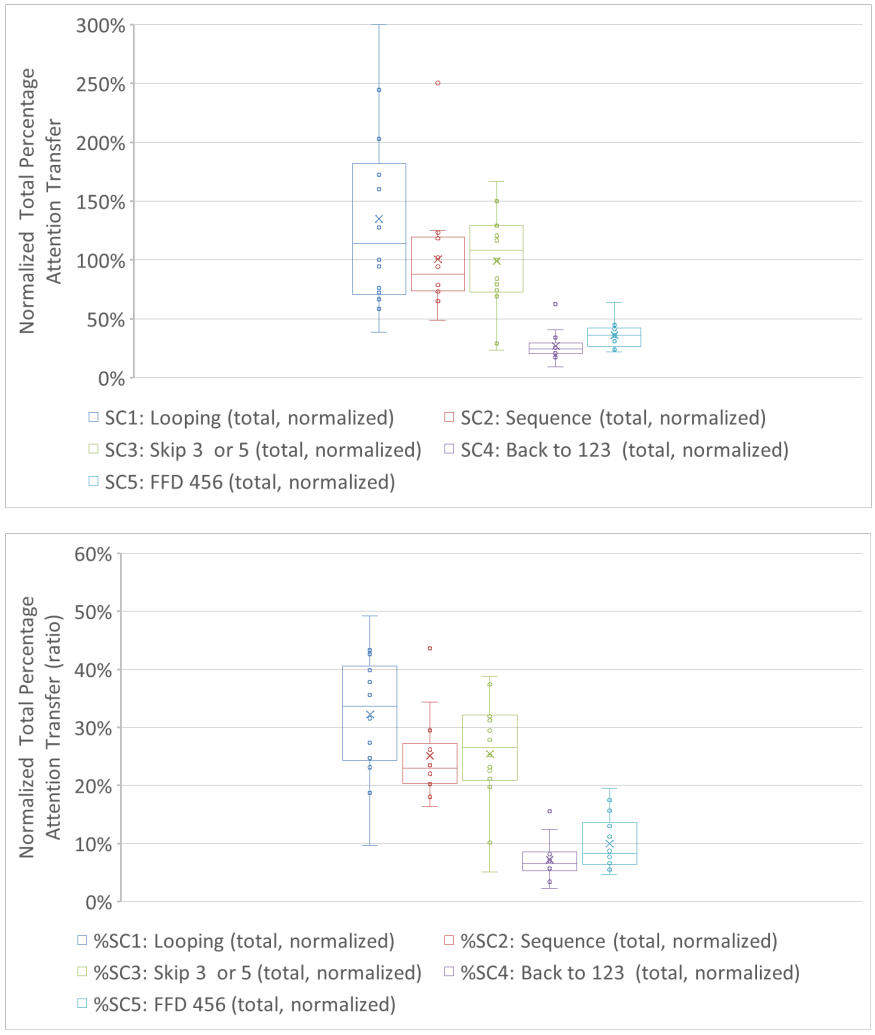


Figure 3.14. Strategy characterization measures box plots – runtime cumulative attention transfer and ratios

Figure 3.15 presents the distribution of sub-characterization measures (total count of transfers and average transfer size) used for a qualitative exploration of characterization modalities. A greater number of transfers for SC1 in Figure 3.15 can be attributed to the fact that it is a repeating action, rather than a one-time action such as SC2. SC1 also shows a greater variability between crews, which may be related to variations in performance. Average attention transfers seem more consistent across the SC measures.

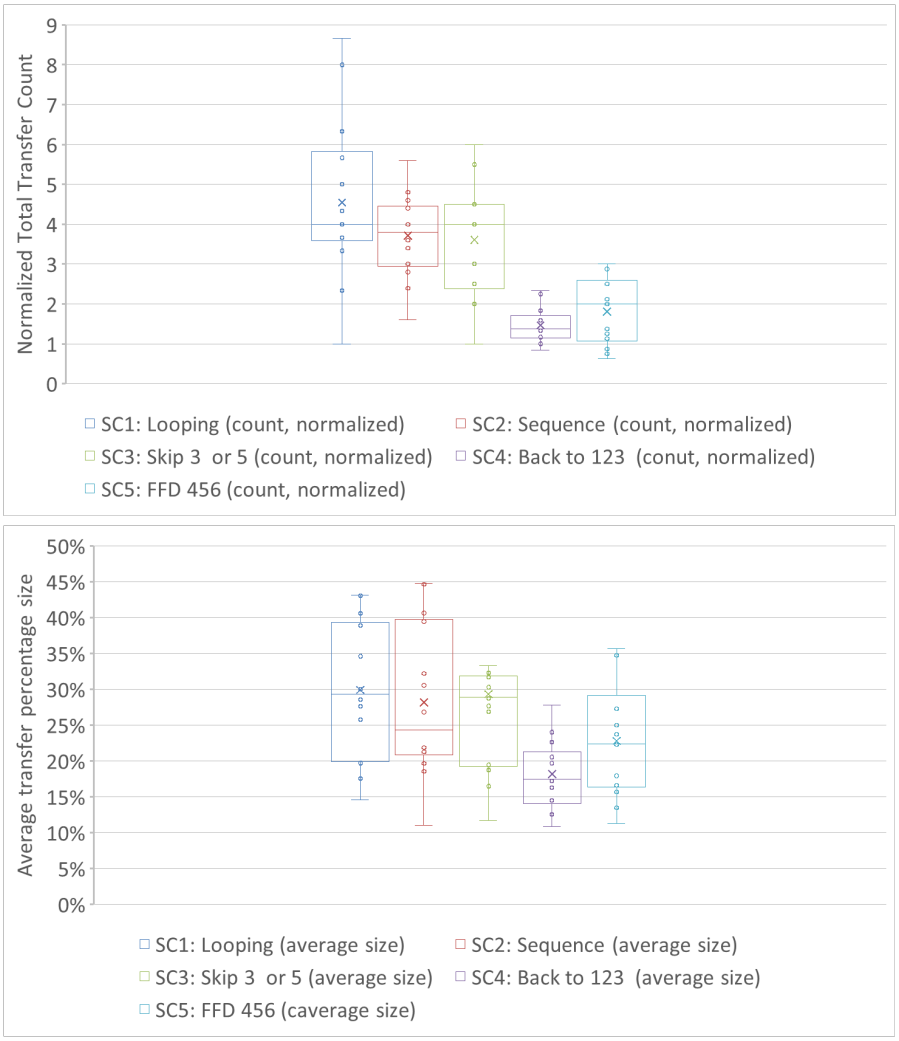


Figure 3.15. Strategy characterization measures box plots – transfer count and average transfer sizes

A review of the characterization modalities (i.e., whether attention transfers were notably large or frequent), did not show clear differences between the top and bottom half of performers (based on Total DFCP) for any of the SC measures, with only SC1 (desirable) possibly indicating that higher performers featured more frequent attention transfers, which would be consistent with increased SC1 cycling behavior. Given these findings, the chapter will proceed to only consider the total transfer variables for the SC measures. The results of the modalities qualitative review can be found in Appendix G.

3.4.3 Behavioral measures results – Principle level

In addition to the strategy phase-level results, Figure 3.16 through Figure 3.18 present the box plots for principle level duration measures (observed, corrected & ratio), principle sequence measures (normalized cumulative total transfers, and ratios) and the sub-characterization measures (transfer count and average attention transfer size). Multitasking correction reduces TM and UM activities considerably, possibly indicating that these activities are practiced more concurrently than CM activities. Analyzing the ratio (percentage) duration metric clearly shows that crews focus on CM more than UM, with also a relative increase in TM (although that metric features greater spread than UM). Crews exhibit (relatively) more PC1 (sequence & looping, desirable) but also some PC3 (Skipping UM, undesirable), which echo's the duration results showing less time spent in UM. Average transfer and transfer count sub-characterizations follow a similar pattern as the total attention measures.

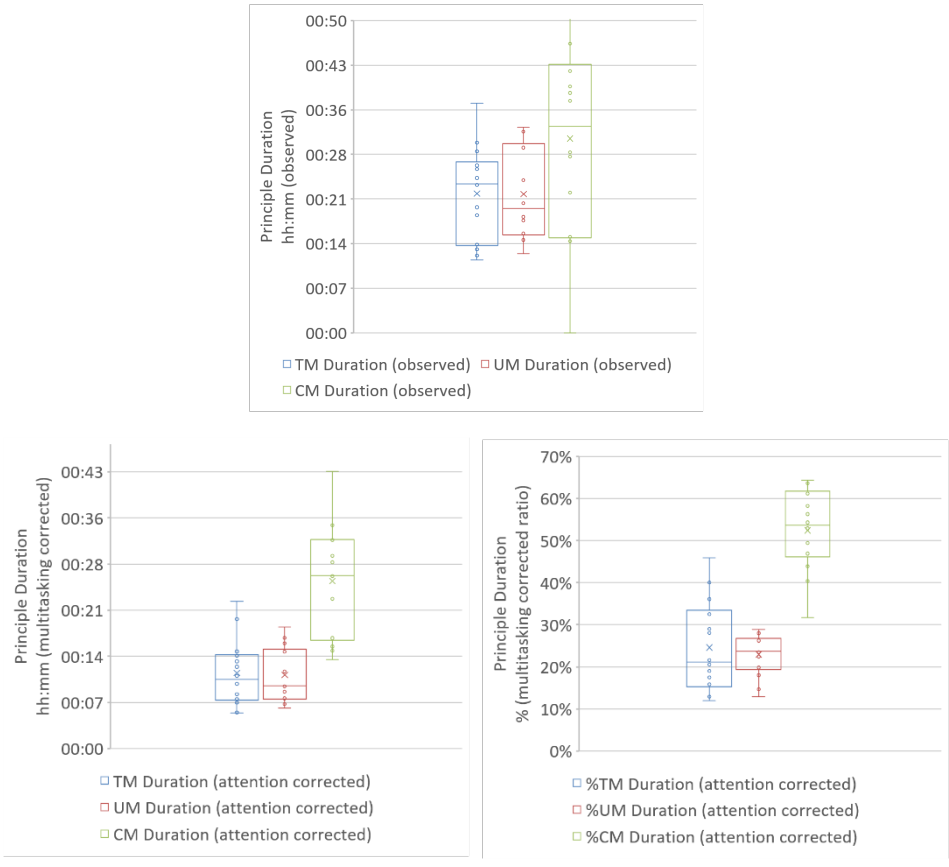


Figure 3.16. Principle duration measure box plots (observed, corrected & ratio)

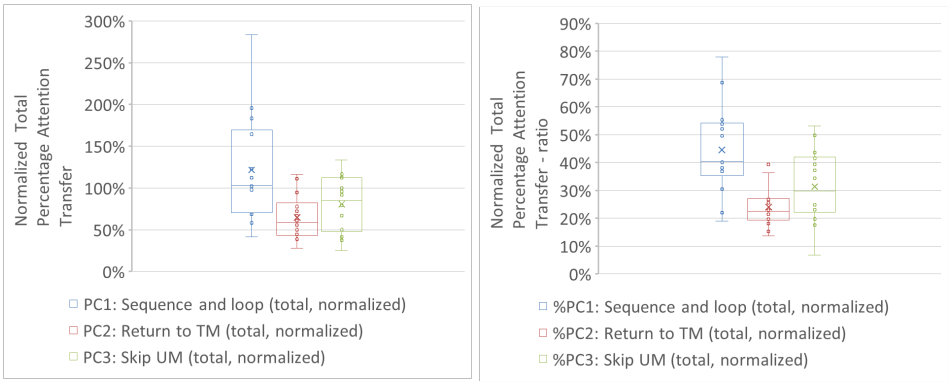


Figure 3.17. Principle sequence measure box plots –runtime cumulative attention transfer and ratios

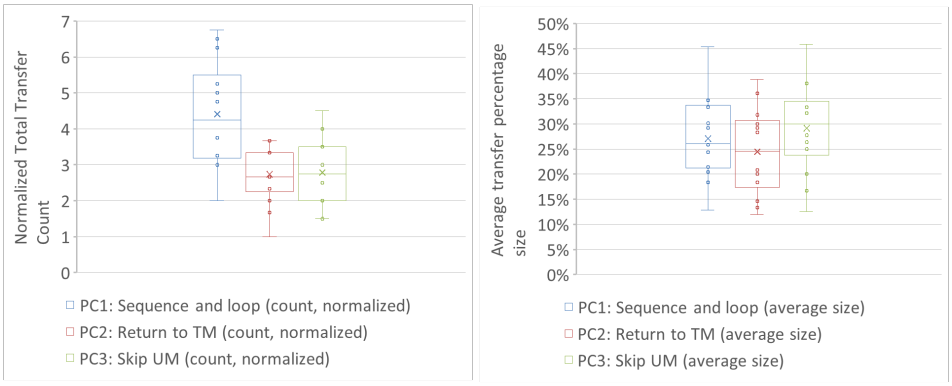


Figure 3.18. Principle sequence measure box plots – transfer count and average attention transfer size

Similar to the procedure phase level analysis, characterization modalities for principle characteristic sequences did not present clear differences between the top and bottom half of performers. Principle level modality results can be found in Appendix G as well.

3.4.4 Behavioral measures correlation – Analysis II

This subsection investigates relations within the behavioral model. As several measures are non-parametric, correlation tests will use Kendall's Tau and with BH *p*-adjustments per measures cluster. Three main correlations are presented in the subsections below: duration versus duration measures, sequence versus sequence measures and duration versus sequence measures. Within each of these correlations, the same set of measures (e.g., six Phase Duration measures) are correlated through different lenses, which provide different insights into behavioral patterns. Duration measures have three correlation lenses (observed; corrected; ratio), characterization measures have two lenses (normalized total sum; ratio) and their cross-correlation will feature six lenses (2x3 lenses). Correlations that persist across

different lenses are particularly strong, which is reported as a robustness indicator indicating how often the correlation is (near) significant. Strategy phase and principle-level measures are kept isolated from each other as they are two different abstraction levels of the same data.

3.4.4.1 Between-duration measures correlation

Duration measures feature three lenses: observed duration (co-occurrence lens), multi-tasking corrected duration (attention lens) and the ratio of corrected durations (pattern lens). Table 3.8 only presents the duration measure correlations which are (nearly) statistically significant (including BH correction separate per lens) within its own matrix/lens, and how they fare in adjacent lenses. Both strategy phase level and principle level correlations are presented (ratios are calculated separate for strategy phase duration measures and principle duration measures). It is expected that latter Phases 4-6 co-exist, but feature inverse time to earlier Phases.

Table 3.8. Overview of between-duration measure correlations (strategy phase and principle level)

Dur. Correlation	Observed τ (p)	Corrected τ (p)	Corrected Ratio τ (p)	Robust-ness	Interpretation
P2-P5	-0.389 (0.055)	-0.495* (0.014)	-0.077 (0.702)	1/3	Trend: more attention provided to P2 relates to less attention to P5
P3-P5	-0.319 (0.112)	-0.275 (0.171)	-0.538* (0.007)	1/3	Trend: P3 and P5 proportions vary inverse with each other
P4-P5	0.407* (0.043)	0.407* (0.043)	-0.143 (0.477)	2/3	Trend: P4 and P5 co-occur and also correlate in attention
P5-P6	0.538* (0.007)	0.187 (0.352)	-0.033 (0.87)	1/3	Trend: P5 and P6 co-occur but after discounting them for multitasking they don't correlate in attention.
TM-UM	0.011 (0.956)	-0.077 (0.702)	-0.385** (0.055)	1/3	After correcting for runtimes with ratio, TM and UM feature a strong negative correlation
TM-CM	-0.143 (0.477)	-0.209 (0.298)	-0.648*** (0.001)	1/3	After correcting for runtimes with ratio, TM and CM feature a strong negative correlation
UM-CM	0.143 (0.477)	0.429*** (0.033)	0.033 (0.87)	1/3	After correcting for multitasking, UM and CM show a positive correlation in attention division

**** Corrected p (q) is significant (q ≤ 0.05) (two-tailed)*

*** Uncorrected p is significant, q is near-significant (p ≤ 0.05; 0.05 < q ≤ 0.1) (two-tailed)*

** Uncorrected p is significant, trend (p ≤ 0.05; q > 0.1) (two-tailed)*

At the strategy-phase level, no significant correlations (passing BH correction) appear but there are trends with moderate ($|\tau| > 0.3$) to strong effect sizes ($|\tau| > 0.5$), which must of course be interpreted with caution. The principle-level analysis provides clearer relations between duration metrics, where UM and CM activities seem to share attention together, and both are negatively related to the proportion of attention spent on TM. These results are supported by the trends at the strategy phase level.

In summary, attention is shared between later phases, but not with initial ones. The inverse may also be true: more time spent in earlier phases reduces time/attention in later phases. Apart from ratio measures [13], the duration values are not zero-sum bounded. Therefore, given the large spread of durations between crews across most phases, this suggests the existence of two distinct crew modalities: 1) being more focused on UM and CM or 2) being more engaged in TM. Correlations with performance should indicate whether this inverse relation has a performance impact.

3.4.4.2 Between-characterization measures correlation

Characterization measures feature two lenses: calculated total runtime sum of attention transfers (normalized to number of possible transfers per characterization) and the ratio between these characterizations (separate for strategy phase and principle-level characterizations). Table 3.9 summarizes (near) significant correlations for both lenses. It is expected that SC1 and SC2 co-exist, but are inversely present to SC3, SC4 and SC5. SC4 and SC5 are expected to co-exist (hypothesized as a low-performance phenomenon).

Table 3.9. Overview of between-characterization measure correlations (strategy phase and principle-level)

Char. Correlation	Total sum τ (*p*)	Char. Ratio τ (*p*)	Robustness	Interpretation
SC1-SC4	-0.221 (0.273)	-0.824*** (0)	1/2	More looping (SC1) is often seen with less returning to P1/2/3 (SC4)
SC1-SC5	-0.044 (0.826)	-0.407* (0.043)	1/2	Trend: More looping (SC1) is often seen with less skipping ahead (SC5)
SC4-SC5	0.144 (0.475)	0.451* (0.025)	1/2	Trend: More skipping ahead (SC5) is often seen with returning to P1/2/3 (SC4)
PC1-PC3	-0.344 (0.089)	-0.736*** (0)	1/2	More sequence & looping (PC1) is often seen with less skipping of UM (PC3)

**** Corrected p (q) is significant (q ≤ 0.05) (two-tailed)*

** Uncorrected p is significant, trend (p ≤ 0.05; q > 0.1) (two-tailed)*

[13] Note: ratio measures do clearly amplify this negative correlation.

Both strategy phase and principle-level correlations indicate that desired sequences (e.g., SC1, PC1) feature a negative correlation to undesired sequences (e.g., SC4, SC5, PC3). While this may seem trivial, this polarization is an important validation of the model where the characteristic measures seem to be complementary and able to model distinct crew behavioral patterns.

3.4.4.3 Duration-characterization measures correlation

The last behavioral correlation is between duration and characterization measures. As behavioral measures are subject to three lenses and characterization measures to two, there are six correlation variations as such, both for the strategy phase and principle-level measures. Table 3.10 presents all (near) significant correlations. It is expected that SC1 and SC2 characterizations co-exist with higher duration in the latter Phases 4-6 and that – conversely – SC3, 4 and 5 co-exist with increased time in earlier phases. These relations may be supported by negative correlations between SC1/SC2 with Phase 1-3 durations, and SC3/SC4/SC5 with Phase 4-6 durations. Similar patterns are expected for principle level variables. As there are several significant correlations, the focus is on whether the correlation is significant (BH corrected) and whether it has a high robustness score. Only results with a robustness score of 3/6 or higher are presented.

Results show clear positive relations between desired behavioral patterns SC1/SC2/PC1 and increase time and attention for uncertainty management and contingency management, and a negative relation to short-term planning time and attention (P3). Undesired patterns such as SC4/PC2 are positively correlated to short-term planning (P3/TM) and negatively related to contingency management. These results confirm that duration and characterization measures co-occur in predictable patterns. Desired behaviors are related to more time in UM and CM activities (which is what those desired behaviors were designed for), while undesired behaviors clearly shift the mode of activity to initial phases. Results are similar at both the strategy and principle level analysis and align with the negative relation between early and later phases described in Subsection 3.4.4.1, and the negative relation between desired and undesired characteristics in Subsection 3.4.4.2.

Table 3.10. Overview of correlation lenses between duration (D-) and characteristics (C-) measures (strategy phase and principle level)

Correlation	D-Observed C-Total sum τ (p)	D-Corrected C-Total sum τ (p)	D-Ratio C-Total sum τ (p)	D-Observed C-Ratio τ (p)	D-Corrected C-Ratio τ (p)	D-Ratio C-Ratio τ (p)	Robustness	Interpretation
SC1-P3	-0.253 (0.208)	-0.319 (0.112)	-0.363 (0.071)	-0.473* (0.019)	-0.451* (0.025)	-0.538** (0.007)	3/6	Trend: Relatively more looping (SC1) is related to less time and attention in short-term planning (P3)
SC4-P3	0.663*** (0.001)	0.685*** (0.001)	0.641*** (0.001)	0.429* (0.033)	0.451* (0.025)	0.538** (0.007)	6/6	More returning to P1/2/3 (SC4) is related to more time and attention in P3
SC1-P4	0.473* (0.019)	0.582*** (0.004)	0.429* (0.033)	0.385 (0.055)	0.407* (0.043)	0.297 (0.139)	4/6	More looping (SC1) is related to more attention to uncertainty management (P4)
SC2-P4	0.575** (0.004)	0.685*** (0.001)	0.663*** (0.001)	0.099 (0.622)	0.033 (0.87)	0.055 (0.784)	3/6	More sequence (SC2) is related to more attention to uncertainty management (P4)
SC1-P5	0.407* (0.043)	0.56*** (0.005)	0.121 (0.547)	0.451* (0.025)	0.560** (0.005)	0.253 (0.208)	4/6	More looping (SC1) is related to more attention to appropriate actions (P5)
SC4-P5	-0.398* (0.048)	-0.309 (0.125)	-0.53** (0.008)	-0.451* (0.025)	-0.560** (0.005)	-0.297 (0.139)	4/6	More returning to P1/2/3 (SC1) is related to relatively less time in appropriate actions (P5)
PC2-TM	0.641*** (0.001)	0.685 (0.001)	0.354*** (0.079)	0.56*** (0.005)	0.604*** (0.003)	0.692*** (0.001)	5/6	More returning to TM (PC2) is related to more time and attention in TM
PC1-UM	0.582** (0.004)	0.846*** (0)	0.538** (0.007)	0.495** (0.014)	0.626*** (0.002)	0.451** (0.025)	6/6	More sequence and looping (PC1) is related to more time and attention in UM
PC2-CM	0.066 (0.742)	-0.088 (0.661)	-0.42* (0.037)	-0.143 (0.477)	-0.429** (0.033)	-0.473** (0.019)	3/6	Relatively more returning to TM (PC2) is related to less attention to CM

*** Corrected p (q) is significant ($q \leq 0.05$) (two-tailed)

** Uncorrected p is significant, q is near-significant ($p \leq 0.05$; $0.05 < q \leq 0.1$) (two-tailed)

* Uncorrected p is significant, trend ($p \leq 0.05$; $q > 0.1$) (two-tailed)

3.4.5 Training effectiveness results – Analysis III

A between group analysis is performed to evaluate whether different treatments (strategy training with/without RIS) changed behavior (hypothesis H_1). It is expected that the STG and DIS crews show more time in latter Phases 4-6, less time in earlier phases, more SC1/SC2 and less SC3/SC4/SC5. All tests will be non-parametric Kruskal Wallis and post-hoc Dunn tests due to small sample sizes for the three groups (BSL N=3, STG N=7, DIS N=4). BH *p*-adjustment is performed within each cluster of measures, as well as for Dunn tests (N=3 groups). Near-significant KW results with strong effect sizes are investigated with post-hoc testing but without BH as these results are only presented as indicative trends. All results are summarized in Table 3.11.

The STG crews spent more time and attention on P2 than other two groups (BSL, DIS), also after correcting for multitasking and runtimes. This may be the result of the STG crews having more awareness of P2 immediate threat management than the untrained BSL crews, while the DIS crews may in turn have had more focus on later phases than the STG crews due to the RIS supporting UM and CM activities. This is counter to the hypothesis that the strategy training would primarily increase time and attention in later phases. The negative correlation between earlier and later phase durations may therefor be the result of the training missing its mark with STG crews, where overfocussing on (or over-structuring) earlier steps reduced time spent on latter phases.

There are non-significant indications that BSL crews spent more time and relative attention on P4/UM (uncertainty management) than STG and DIS crews, and that the BSL crews skipped ahead more (SC3) than the DIS crews. This could be attributed to a less structured approach for UM in BSL crews. In contrast, the STG and DIS groups did not spend more time in UM, which also suggests the training may not have been as effective as thought.

The most profound group effect discovered was that the DIS group showed much more observed behavior in P5 and P6 than the STG group, possibly due to the addition of the RIS. However, this effect isn't present when correcting for multitasking, which may indicate that while DIS crews have been observed performing P5 and P6 activities more often, they haven't actually allocated more attention to them (i.e., other concurrent tasks are discounting this). That may be the benefit of the display: it offloads some of the P5 and P6 tasks and allows for a wider distribution of cognitive capacity (i.e., to P4 uncertainty diagnostics). In summary, while the results cannot confirm an effective training effect, there are indications that the RIS may have supported more effective management of ambiguity and opacity.

Table 3.11. Significant and near-significant training effects (strategy phase and principle levels combined)

Cluster & measure	KW effect	KW sig (BH q)	Dunn pairs	Dunn effect	Dunn sig p (BH q)	Median comparison & interpretation
C1 P2 Duration (observed)	0.582***	<0.001 (<0.001)	BSL-STG	0.897***	0.005 (0.015)	STG shows 03m00s more time in P2 manage immediate threats than BSL, and 02m50s more than DIS. BSL and DIS have the same observed P2 durations.
			BSL-DIS	0.193	0.522 (0.522)	
			STG-DIS	0.206***	0.019 (0.029)	
C1 P4 Duration (observed)	0.203Δ	0.083	BSL-STG	0.642*	0.042	Strong KW/Dunn effect: BSL shows +08m46s more time in P4 uncertainty management than STG.
			BSL-DIS	0.189	0.531	
			STG-DIS	0.444	0.141	
C1 P5 Duration (observed)	0.413***	0.008 (0.024)	BSL-STG	0.343	0.255 (0.255)	DIS median shows 20m16s more time in P5 appropriate actions than STG.
			BSL-DIS	0.377	0.211 (0.317)	
			STG-DIS	0.879***	0.005 (0.015)	
C1 P6 Duration (observed)	0.407***	0.009 (0.018)	BSL-STG	0.209	0.488 (0.488)	DIS median shows 18m59s more activity in P6 long-term planning than STG.
			BSL-DIS	0.495	0.1 (0.15)	
			STG-DIS	0.874***	0.006 (0.018)	
C2 P2 Duration (corrected)	0.567***	0.001 (0.006)	BSL-STG	0.662***	0.028 (0.042)	STG shows 00m41s more attention on P2 immediate threats than BSL, and 01m01s more than DIS. DIS and BSL have same P2 duration.
			BSL-DIS	0.11	0.715 (0.715)	
			STG-DIS	0.904***	0.004 (0.012)	
C3 P2 Duration (%corrected)	0.567***	0.001 (0.006)	BSL-STG	0.662***	0.028 (0.042)	STG shows 1.52% more relative attention on P2 manage immediate threats than BSL, and 1.95% more than DIS. BSL and DIS have same P2 duration ratios.
			BSL-DIS	0.11	0.715 (0.715)	
			STG-DIS	0.904***	0.004 (0.012)	
C3 P4 Duration (%corrected)	0.244Δ	0.057	BSL-STG	0.587Δ	0.052	Strong KW/Dunn effect: BSL shows 4.81% more relative attention on P4 uncertainty management than STG, and 7.72% more than DIS. STG and DIS have the same P4 duration ratios.
			BSL-DIS	0.701*	0.027	
			STG-DIS	0.168	0.577	
C4 SC3: Skip P3 or P5 (undesirable)	0.199Δ	0.067	BSL-STG	0.403	0.181	Strong KW/Dunn effect: BSL median shows 55% more total attention transfers skipping ahead to P4-6 than DIS.
			BSL-DIS	0.718Δ	0.069	
			STG-DIS	0.39	0.195	
C6 UM Duration (observed)	0.203Δ	0.083	BSL-STG	0.642*	0.042	Strong KW/Dunn effect: BSL median shows 8m46s more time in uncertainty management than STG.
			BSL-RIS	0.189	0.531	
			STG-RIS	0.444	0.141	

*** Corrected p (q) is significant ($q \leq 0.05$) (two-tailed)

** Uncorrected p is significant, q is near-significant ($p \leq 0.05$; $0.05 < q \leq 0.1$) (two-tailed)

* Uncorrected p is significant, trend ($p \leq 0.05$; $q > 0.1$) (two-tailed)

Δ Strong effect size, Uncorrected p is not significant, trend ($0.05 < p \leq 0.1$) (two-tailed)

3.4.6 Behavior versus performance correlation – Analysis IV

This section will investigate whether specific behaviors relate to improved DFCP performance (hypothesis H_2) by evaluating the entire study population in a within-group analysis at both the strategy phase and principle level. Table 3.12 summarizes all (near-)significant correlations between all behavioral measures and the five DFCP performance indicators. While several correlation results feature a significant *p*, post-BH *p*-adjustments indicate no significant correlations after correcting for false discovery rates. Those results that featured a significant *p* are still presented and discussed for general trends.

In terms of duration measures, more time and attention in short-term planning (P3) is related to lower Route DFCP scores. Conversely, more time and attention toward appropriate actions (P5) is related to higher Route DFCP scores. This also reflects the inverse relationship between P3 and P5 durations and suggests that improved route management requires taking more time for and bringing attention to contingency management (i.e., future threats, also related to routing) rather than spending a lot of time and attention on near-term (possibly reactive) routing decisions. This result is consistent at the principle level which shows a negative relation between TM and Route DFCP, and inversely a positive relation between CM and Route DFCP. Other relations with duration measures indicated that both Total DFCP and ENG1 DFCP performance benefited from more activity in long-term planning (P6). Particularly for the ENG1 DFCP, this is relevant as it presents considerable energy management considerations with an unreliable autothrust system and binary engine control alternatives. Finally, ENG2 DFCP improves with more attention towards P1 stabilize flight path. This is reasonable, considering the impact on the flight path to mitigate and manage ENG1 overheat issues.

In terms of behavioral characteristics, only Route DFCP scores featured correlations with these measures. SC1 (looping P4/5/6) related to increased Route DFCP scores, possibly reflecting a re-appraisal of route options as they appear or close out. Interestingly, skipping to these later phases (SC4, PC3) correlated negatively with Route DFCP. In other words, time spent in and cycling of later phases (P4/5/6) increases Route DFCP but does require some attention in earlier phases (albeit limited, looking at the negative relation between P3 and Route DFCP). This may prove to be an indication that some sequencing is valuable to stabilize the flight and buy time (and cognitive capacity) to engage in the sensemaking done in the later phases.

Table 3.12. (Near-)significant correlations between behavior and performance (strategy phase and principle levels combined)

Cluster & Behavioral measure	DFCP Measure	Kendall τ	Sig *p* (BH *q*)	Interpretation
C1 P3 Duration (observed)	Route DFCP	-0.626*	0.004 (>0.1)	Trend: More time spent on short-term planning P3 worsened route management
C1 P5 Duration (observed)	Route DFCP	0.476*	0.028 (>0.1)	Trend: More time spent on appropriate actions P5 improved route management
C1 P6 Duration (observed)	Total DFCP	0.412*	0.046 (>0.1)	Trend: More time spent on long-term planning P5 improved overall performance
C1 P6 Duration (observed)	ENG1 DFCP	0.467*	0.031 (>0.1)	Trend: More time spend in long-term planning P6 improved ENG1 management
C2 P1 Duration (corrected)	ENG2 DFCP	0.418*	0.049 (>0.1)	Trend: More attention to flight path stabilization P1 improved ENG2 management
C2 P3 Duration (corrected)	Route DFCP	-0.551*	0.011 (>0.1)	Trend: More attention to short-term planning P3 worsened route management
C2 P5 Duration (corrected)	Route DFCP	0.451*	0.037 (>0.1)	Trend: More attention to appropriate actions P5 improved route management
C3 P3 Duration (%corrected)	Route DFCP	-0.626*	0.004 (>0.1)	Trend: Greater proportion of attention toward short-term planning P3 worsened route management
C3 P5 Duration (%corrected)	Route DFCP	0.426*	0.049 (>0.1)	Trend: Greater proportion of attention toward appropriate actions P5 improved route management
C4 SC1: Looping (desired)	Route DFCP	0.476*	0.028 (>0.1)	Trend: More looping of P4/5/6 improved route management
C5 SC1: Looping (%) (desired)	Route DFCP	0.551*	0.011 (>0.1)	Trend: More looping of P4/5/6 improved route management
C5 SC4: FFD 456 (%) (undesired)	Route DFCP	-0.476*	0.028 (>0.1)	Trend: Skipping ahead to P4/5/6 worsened route management
C6 TM Duration (observed)	Route DFCP	-0.426*	0.049 (>0.1)	Trend: More time spent on TM worsened route management
C7 TM Duration (%corrected)	Route DFCP	-0.501*	0.020 (>0.1)	Trend: Greater proportion of attention toward TM worsened route management
C7 CM Duration (%corrected)	Route DFCP	0.526*	0.015 (>0.1)	Trend: Greater proportion of attention toward CM improved route management
C10 PC3: Skip UM (%) (undesired)	Route DFCP	-0.426*	0.049 (>0.1)	Trend: Skipping UM worsened route management

* Uncorrected *p* is significant, trend ($p \leq 0.05$; $q > 0.1$) (two-tailed)

3.5 Reflection on hypotheses

The reflection is limited to the two hypotheses H_1 and H_2, at both the strategy phase and principle levels as these draw conclusions at different levels of abstraction. The implications of the findings for the operational and research domains are covered in Section 3.6.

3.5.1 Training effectiveness conclusions

Table 3.13 summarizes the response to hypothesis H_1, segregating key behavioral expectations.

Table 3.13. Overview of study conclusions for hypothesis H_1

Training in strategy/RIS will …	Conclusion
increase time & attention in Phases 4, 5 and 6.	Reject. The DIS group (also trained in the strategy) did show more activity in P5 and P6, but this was gone after correcting for multitasking. The RIS may have been causal to this shift. Furthermore, there are weak indications that the BSL may have spent more time in P4.
reduce time & attention in Phases 1, 2 and 3.	Reject. The training actually (slightly) increased time & attention in P2 for the STG group.
increase attention toward SC1 and SC2 sequences.	Reject. There are no indications that training increased SC1 and SC2.
reduce attention toward SC3 (mild effect), SC4 and SC5 sequences.	Partial accept. There are weak indications that untrained crews showed more SC3. Training effects for SC4 and SC5 cannot be confirmed.
increase time & attention in UM and CM.	Reject. There are no indications that training increased attention in UM and there are weak indications that the BSL may have spent more time in UM.
reduce time & attention in TM.	Reject. There are no indications that training reduced attention in TM.
increase attention towards PC1 sequences.	Reject. There are no indications that training increased PC1.
decrease attention towards PC2 and PC3 sequences.	Reject. There are no indications that training decreased PC2 and PC3.

The above findings indicate that the training was not effective to induce the behaviors that were hypothesized to have been more prevalent after training. There are many reasons for this, including the brevity of the training (only several hours of classroom and simulator familiarization before the actual experiment flight). Another explanation may be that making crews consciously aware of their actions may result in them re-navigating certain phases and steps, which may be the reason for the STG group's increased time and attention

for Phase 2. This was observed on some occasions, where crews would be taking actions and then, mid-way through their actions, remember they had a strategy to apply, which caused them to retrace some of their steps. This would be attributable to insufficient time to truly internalize and familiarize with a new way of managing complex situations. Another reason for these results may lie in random effects due to the small sample sizes. If the STG and DIS groups were coincidentally provided crews that struggled more, and the BSL group featuring crews naturally proficient at managing complexity, then this would also limit the power of these between-group results. An experiment with the same group flying two flights – one before training and one after – may result in a better pairwise training effect study.

3.5.2 Behavior-performance conclusions

Table 3.14 summarizes the response to hypothesis H_2, segregating key expectations where behaviors improve performance.

Table 3.14. Overview of study conclusions for hypothesis H_2

Performance will increase with …	**Conclusion**
increased time & attention in Phases 4, 5 and 6.	Partial accept. There are trends that time and attention in P5 related positively to Route DFCP. Time in P6 was also related positively with Total and ENG1 DFCP scores.
reduced time & attention in Phases 1, 2 and 3.	Mixed results. Strong trends that time & attention in Phase 3 related negative with Route DFCP scores. At the same time, a trend indicates that attention in P1 relates positively with ENG2 DFCP scores.
increased attention toward SC1 and SC2 sequences.	Partial accept. Trend that SC1 related positively to Route DFCP scores.
reduced attention toward SC3 (mild effect), SC4 and SC5 sequences.	Partial accept. Trend that SC4 related negatively to Route DFCP scores.
increased time & attention in UM and CM.	Partial accept. Trends indicate relative attention to CM related positive to Route DFCP scores.
reduced time & attention in TM.	Partial accept. Strong trends indicate more time and attention in TM related negatively to Route DFCP scores.
increased attention towards PC1 sequences.	Reject. There are no indications that PC1 increases performance.
decreased attention towards PC2 and PC3 sequences.	Partial accept. There are indications that PC3 is negatively related to Route DFCP scores.

In contrast to the training effect results, there are more consistent relationships between behavior and performance measures. It should be reiterated that all behavior-performance correlations only provided trends (i.e., significant *p*-values but non-significant

q-values). However, given relatively strong effect size (~ $\tau \pm 0.5$), confirmation of these relations may be possible with a stronger testing power (e.g., with more participants).

The observed trends do suggest that performance improves by not skipping phases (PC3) or going back (SC4), while also improving when focusing time and attention on later phases (Phases 4, 5 and 6) related to UM and CM, looping them (SC1) and spending less time in earlier phases related to TM. These results were particularly clear when considering performance measures related to managing future threats and contingencies (e.g., routing and energy management for this scenario). While not all behavioral measures were clearly correlated to performance, the strong correlations *between* the behavioral measures themselves may provide further indications about their effect on performance. For example, the strong *negative* correlation between PC1 and PC3 behavioral measures, and *negative* relation between PC3 and performance, may suggest a *positive* relation between PC1 and performance. Such proxy trends form specific hypotheses for future testing and require greater testing power to be confirmed.

In summary, the study was able to confirm that certain behaviors (brief TM, avoiding skipping/going back, UM and CM focus with looping) are possible performance drivers in the context of dealing with opacity and ambiguity on the flight deck. More powerful testing (possibly a meta-analysis) is required to confirm these trends beyond the False Discovery Rate (FDR). Furthermore, given the limited experimental data, the study concludes that the proposed treatment (strategy training approach, and RIS) cannot be proven effective in inducing desired behavior. The indications that the RIS positively changed behavior are stronger, as evident by more observed behavior related contingency management in DIS crews (which was the purpose of the RIS). It may be that by illustrating more dimensions of risk, the RIS facilitates a broader threat assessment by the crew and identify better flight continuation options. Both improvements in the training, as well as improvements in the study design (increased power and group sizes) may yield more encouraging results in identifying what training or system developments may instill the desired behaviors.

3.6 Discussion

Reflecting on the above study, there is some evidence that certain intended behaviors such as effective (i.e., brief) short-term stabilization and planning to buy time, and subsequently using this time to managing uncertainties and contingencies, may increase crew performance in dealing with complex, ambiguous and opaque flight situations. This evidence supports the fundamental shift from an *executive* to a *learning* function or mindset that the crew must engage to maintain safety performance when faced with opacity and ambiguity as in this study's scenario. This requires time and space for sensemaking, an explorative mindset

("managing uncertainties") and also a feedback process to reflect on the new knowledge gained through interacting with the situation and systems. A similar conclusion is drawn by Woltjer et al. (2015/2026) who examined Man4Gen study results through the ECOM lens (from Chapter 2) and also underscores the value of short-term planning and subsequent progressing to long-term planning. The authors also supported this ECOM analysis with a similar behavioral mapping process as described in this chapter.

From an operational perspective, the results (pending further analysis) may drive a change in how flight crew training and flight decks can be reconfigured to be more effective in the context of a crew dealing with complexity, ambiguity and opacity which are becoming the Achilles' heel of modern airliners (as well as other complex systems in- and outside of aviation). While this study lacked the statistical power to confirm the maturity of the strategy and RIS in their current form, results are consistent and show positive effects which warrant revisiting them with a larger study population.

While the study's small population size limits its power to evaluate training and RIS effectiveness, qualitative observations indicate that the strategy, and how it was trained, may be subject to future improvements. While some STG/DIS crews easily integrated the new strategy into their workflow, most crews were observed recalling the strategy somewhere midway through the scenario, rather than engaging the strategy when the initial problems occurred (i.e., the ENG2 overheat or autothrust system failure). Longer training time (possibly months in advance) and more practice will likely improve strategy recall and application consistency. Improved training may also include more case studies to teach and explore a more dynamic response centered on the core principles of TM → UM → CM, rather than the rigidity that a quick reference card introduces. Despite the limited training, feedback from the STG group about the strategy does indicate that they deemed it as helpful, most notably by well-performing crews (based on Total DFCP scores).

Reflecting on the experimental setup and approach, this study could have benefited from a less complicated experimental logistics which undoubtedly contributed to data noise (i.e., two different locations, two different research simulators, two aircraft classes, two different airline participant pools, two research teams and two different experiment executions with and without RIS). Although great attention has been given to provide equivalent scenarios from a challenge-centric point of view (Niedermaier et al., 2018) and to align experiment execution, debriefing and scoring mechanisms, some noise in the data will be attributable to this somewhat divided setup. Reproducing the study with a more homogenous sample size and setup may reduce random effects and improve statistical power. This may also be done by means of a longitudinal study design evaluating crew before and after introducing the strategy and RIS to reduce group effects and correct for differences

in crew natural abilities. Finally, future (meta-studies) may also investigate improved behavioral measures (e.g., compound measures that cluster desired and undesired patterns) as well as new measures such as looking at the temporal distribution of phase activities and characterizations (not just the runtime sums). Despite these limitations, the analysis presented in this chapter aims to catalyze industry insight into the minutiae of flight crew sensemaking in the face of opacity and ambiguity on the flight deck, where both the results as well as innovative behavioral mapping methods invite new ways of looking—a new perspective—toward flight crew performance in modern transport aircraft.

DISCLAIMER *The data and results were produced as part of the FP7 2012 Aeronautics and Air Transport program under EC contract ACP2-GA-2012-314765-Man4Gen. The NLR and DLR are owners of the data and results of this research and are published with permission from both NLR and DLR. The views and opinions expressed in this publication are those of the authors and are not intended to represent the position and opinions of the Man4Gen consortium and/or any of the individual partner organizations.*

4
The Airmanship 2.0 concept

Following the retrospective results from Chapters 2 and 3, Chapter 4 proceeds to address the challenge of integrating behaviors effective in opaque operations, with existing, determinist behavioral standards from the currently predominant transparent operations paradigm. The new compound, hybrid pilot model is contrasted against several notable aviation accidents in the past two decades.

The content of this chapter is based on:

Paper title	Solving the brittleness problem: Redefining airmanship in the age of increasing complexity
Author(s)	Mohrmann, J.F.W.
Published in	Pre-print (engrXiv, 2026)
DOI	https://doi.org/10.31224/7310

***ABSTRACT**. The increasing performance of our aviation system has in-part been made possible through growing complexity of systems and operations, together with a drive toward standardization and compliance as safety foundations. Despite high levels of safety performance and reliability, commercial aviation accidents in the past two decades reveal a systemic, brittle crew response when faced with opacity and ambiguity resulting from this complexity. Building on research into effective flight crew behaviors and strategies in complex fourth generation commercial transport aircraft presented in Chapters 2 and 3, this chapter proposes a new concept for the human pilot operator – Airmanship 2.0 – to address the brittleness of the existing socio-technical flight deck system and further improve aviation safety. This concept proposes a compound pilot role to systematically harmonize compliant and adaptive behaviors and mitigate the current bifurcation of and related strain on the pilot role. This chapter contrasts the proposed Airmanship 2.0 concept against notable accidents in the past decades.*

4.1 Introduction

Per 2019 the worldwide commercial air transport system transported approximately 4.5 billion passengers with an accident rate of only 1.13 per million flights (IATA, 2020). Compared to the year 2001, 2019 featured a 235% increase in number of sectors flown, while also achieving 60% reduction in the number of accidents per one million flights (IATA, 2002; IATA, 2020). However, as the aviation industry's safety performance is approaching a 10^{-7} accident rate, it is showing signs of stabilizing into a non-plus ultra-safe system of the existing system design (Stoop, 2017; Airbus, 2025).

Safety studies into accidents and incidents from the past two decades (IATA, 2014b; Man4Gen Consortium, 2015; Mohrmann et al., 2015) reveal specific flight deck dynamics in modern aircraft as a key limiting factor of system safety performance. This is the result of increasing cognitive distance between modern crews' and aircraft systems and operations, suffering from both the growing complexity of modern aircraft and operations, as well as a decrease in the requisite crew capacity to respond adaptively to such complexity. This cognitive distance is not a wholly novel phenomenon: Mosier et al. (1998) and Manzey et al. (2012) described similar performance-limiting mechanisms such as automation bias and automation complacency. However, in recent years these mechanisms are becoming the major barriers to improving safety, and may be exacerbated by future increase of complex and potentially even more opaque systems such as Human-Artificial Intelligence (AI) teaming (HAT) (Eurocontrol, 2020; EASA, 2023; Lorrig & Daw, 2024; Kirwan 2025; Hörhager, 2025) as well as the advent of Reduced Crew Operations (RCO) (Malick &

Gollnick, 2016; Sprengart et al., 2018; Schmidt & Stanton, 2020), or their combination (Minaskan et al., 2022).

Research into this control gap has been able to reproduce flight crews' difficulty to manage opaque and ambiguous situations in the flight deck (Sarter & Woods, 1994; Saurin & Carim, 2012; IATA, 2014b; Man4Gen Consortium, 2015; Mohrmann et al., 2015; Stoop & Van Kleef, 2015; Wolter et al., 2015; Strauch, 2017; Kharoufah et al., 2018; Proctor & Van Zandt, 2018; Kelly & Efthymiou, 2019; Prinzel et al., 2024; Gago et al., 2025), as well as revealing flight crew responses that *are* effective in managing complexity and related opacity (Rankin et al., 2016; Field et at., 2017; Mohrmann et al., 2017; FutureSky P6 Consortium, 2017; Field et al., 2018; Banks et al., 2020; Landman et al., 2020; Hancock et al., 2022; Sarter, 2024). Behavioral studies such as Man4Gen as well as industry initiatives such as Competency Based Training and Assessment (CBTA) and Evidence Based Training (EBT) (ICAO, 2013; EASA, 2020; IATA, 2024b) acknowledge the need to adjust the pilot role toward the inclusion of more effective responses to further improve flight safety.

The dynamic nature of such responses, however, lies in stark contrast to the long-held (and still robust) performance standards centered on compliance, error-reduction and reproducibility which account for the strong safety performance in recent decades. The contrast between and need for both compliant and adaptive control paradigms strains the pilot role. This chapter explores this dichotomy of compliant versus adaptive control and sets forth a new, compound concept for the pilot role, *Airmanship 2.0*, as an effective, efficient and responsible assimilation of both control modes. The concept is explored in the context of seven commercial aviation accidents which occurred in the past two decades.

4.2 The current flight deck control paradigm

The proposal of a new framework for the pilot model is incumbent on several concurrent developments in aircraft systems, the pilot profession and the interaction between them.

4.2.1 Evolution of aircraft systems

The past decades have seen significant evolutions in flight deck systems across four generations of flight decks (Abbott, 2017; Airbus, 2025). To meet increasing operational, environmental and economic demands, flight deck systems feature high levels of integration and level of autonomy (Boy, 2020; Clark & Wilson, 2024). Key examples of such systems include flight envelope protection and active upset recovery technology, GPS-based precision navigation, automatic center-of-gravity fuel-balancing systems, automated engine startup and the advent of paperless flight decks.

The growing number of systems also drives more system integration such as EICAS/ECAM [14] fault management systems which directly invoke Quick Reference Handbook (QRH) checklists to be executed. Similarly, digital checklists can detect the state of several aircraft system configurations and monitor checklist completion (e.g., the "No Blue" callout during Airbus approaches) (Landry, 2009). The Flight Management System (FMS) can receive a flight plan from an operator's dispatch team via a wireless connection at the gate or in the air via CPDLC [15], which automatically does performance calculations for weight and balance, take-off performance, flight navigation and auto-flight system configurations, all with minimal crew interactions (mostly confirmations). Bowles (2018) and Feary (2018) provide an overview of this evolution in automation, including an interesting perspective of *Automation Readiness Levels*, for further reading.

A growth in system autonomy is also seen in flight envelope protection systems (Landry, 2009), Primary Flight Display (PFD) declutter modes (SAIA, 2016), automated ground-based support systems (Ho et al., 2017) and future free-flight operations with aircraft autonomously interacting to manage traffic flows (Enea & Porretta, 2012). The developments and promises of AI may further expand the autonomy of systems (Kulida & Lebedev, 2020; Lorrig & Daw, 2024), also echoed in the AI roadmaps provided by EASA and the FAA (EASA, 2023; FAA, 2024).

4.2.2 Evolution of the pilot profession

Concurrently, the pilot role has shifted. Boeing predicts 804,000 new pilots (Boeing, 2019) required in the next two decades, while average pilot salaries have dropped to less than 20,000 USD in the past decade (Crouch, 2020) together with reduced job security (Fraher, 2019). This trend is very present in the US where the first 1,500 flight hours requires hour-building in at lower-pay regional operators (Caraway, 2020). In contrast, the financial burden of flight training has increased with higher fuel prices and state- or airline-funded training programs being replaced with loans and even pay-to-fly schemes (Valenta, 2018; Maxwell & Grant, 2021).

In addition to this, pilot training efforts have been under pressure to answer for pilot shortages by increasing throughput, as well as remaining commercially viable in a competitive, low-margin market (Soo, 2018; Adanov et al., 2020). On-aircraft familiarization has made way for more simulator and line training such as Zero Flight Time Training (ZFTT), ab-initio training has made way for Multi Pilot License (MPL) training (Wikander & Dahlström, 2016), reducing (frozen) Air Transport Pilot License (ATPL) small aircraft

[14] EICAS: Engine Indication and Crew Alerting System; ECAM: Electronic Centralized Aircraft Monitor

[15] CPDLC: Controller Pilot Data Link Communications

flying time from 200 to 120 hours and condition cadets specifically towards the right seat of a regional commercial aircraft such as an A320 or 737.

Lastly, the (r)evolution in airline networks is negatively affecting pilot fitness. Twelve- to fifteen-hour flight routes and 24-hour airport operations have extended work into non-circadian times (Sallinen et al., 2020; Shah, 2024). This is acknowledged by efforts which put considerable effort into recognizing and mitigating fatigue risks for example through the implementation of Fatigue Risk Management Systems (FRMS) (ICAO, 2016; UK CAA, 2019). The combination of financial stress, circadian stress and other demands such as cultural diversity in the flight deck, forced relocations and the strong cyclic nature of the industry puts increasing strain on the physical, mental and emotional capacity of the human pilot (Cullen et al., 2021).

4.2.3 Evolution of pilot-system interactions

The above developments in technology, training and operations describe a shift in flight deck collaboration (Landry, 2009; Abbot, 2017; Boy, 2020; Clark & Wilson, 2024). The ten levels of automation listed in Table 4.1 serve as a useful reference to describe this shift (adapted from Parasuraman et al., 2000). "Implementation" tasks were the first to be offloaded to automation around the 1950's with the advent of autopilot systems (around levels 3-4), "Generate" tasks shifted to automation since the 1960's with the introduction of warning systems (around levels 5-6), "Select" shifted to automation since the 1990's with the introduction of EICAS and ECAM systems (around levels 7-8), leaving "Monitor" for the pilot in the most recent years (currently around levels 8-9).

Table 4.1. Levels of Automation (Parasuraman et al., 2000)

	Information-Processing Functions H: Human, C: Computer			
Levels of Automation	**Monitor**	**Generate**	**Select**	**Implement**
Manual control	H	H	H	H
Action support	H/C	H	H	H/C
Batch processing	H/C	H	H	C
Shared control	H/C	H/C	H	H/C
Decision support	H/C	H/C	H	C
Blended decision making	H/C	H/C	H/C	C
Rigid system	H/C	C	H	C
Automated decision making	H/C	H/C	C	C
Supervisory control	H/C	C	C	C
Full automation	C	C	C	C

This evolution can be described as a shift from a pilot-centric model (where the automation supports the human) to an automation-centric model (where the human monitors the automation), shepherded through key human-machine design philosophies such as human-centered design (Kirwan, 2025), cognitive ergonomics (Harris, 2017) and human-systems integration (Boy, 2020). More recently, research and policy making in human-AI teaming (HAT) also anticipate a similar transition toward greater autonomy and a reduced role for the human operator (EASA, 2023; Lorrig & Daw, 2024; Kirwan, 2025). An example of this can be found in the EU HAIKU project which defines an evolution through six levels of HAT from enhancing information for human operators through to full AI autonomy (Kirwan, 2025).

These developments in autonomy and system integration cognitively isolated the flight crew from control loops and, combined with the highly reliable nature of systems, induce monitoring errors and automation bias (Mosier et al., 1998; Mumaw et al., 2001; Manzey et al., 2012; Bradshaw et al., 2013; Man4Gen Consortium, 2012; Schutte, 2015; Man4Gen Consortium, 2015; Sarter, 2024). Ironically, the mitigation of such new human error modes furthers the shift toward automation in the flight deck. This process of automation deference was already described by Bainbridge (1983). Klein (2011) aptly describes this process as the *Self-Reinforcing Complexity Loop*. Kirwan (2025) also highlights the risk of future AI systems exacerbating the already-present opacity of automated systems and suggests that existing human factors design methodologies may not be sufficient for such future teaming modalities. Kox et al. (2026) provides an interesting exploration into HAT by observing specific AI applications in the cockpit through existing CRM and Big 5 Teamwork (B5T) frameworks. While this plays toward responsible design, longitudinal studies will be required to truly understand how the introduction of AI influences cockpit dynamics, as undesired socio-technical system responses may only emerge as a result of progressive skill and knowledge atrophy, and commensurate drift to human overreliance on (new) automation.

4.2.4 The brittleness problem

The strong interactions between the above developments in systems, the pilot role and operations require appreciating their collective performance as a Joint Cognitive System (JCS) (Hollnagel & Woods, 2005; Woods, 2019). It should be noted that current flight decks have largely been designed from a traditional human–machine perspective, in which pilot and technical system are treated as separate units of analysis connected through an interface, which is currently not considered an JCS as such. The combination of rigid (deterministic) automated systems, flight crews strongly conditioned toward procedural compliance and rationed information-sharing between systems and pilots results in a Flight Deck Socio-

Technical System (FDSTS) that is performing very strongly within the domain of deterministic (known) operations.

However, as the accident cases in Section 4.5 will illustrate, beyond the realm of deterministic operations, when faced with opacity, the current FDSTS struggles to respond in an adaptive way and make sense of "unscripted" situations. The strongly conditioned rigidity of both systems and pilots in the FDSTS sacrifices cognitive flexibility, dynamic responses and sensemaking (Manzey et al., 2012; Man4Gen Consortium, 2012; Mohrmann et al., 2015; Mohrmann et al., 2017; Sarter, 2024). As such, the FDSTS can better be described as *brittle*: featuring high performance through reproducibility within the deterministic operational domain, yet catastrophic failure (to adapt) beyond that domain (Woods, 2015; Hancock et al., 2022).

Figure 4.1 organizes several factors that drive this brittle characteristic. These factors have been reproduced in several research activities such as the EU FP7 project Man4Gen (Man4Gen Consortium, 2015), a UK CAA investigation into pilot fatigued performance (UK CAA, 2019), an EU-contracted global pilot fatigue data study (Sallinen et al., 2020), an EASA-contracted startle and surprise management study (Field et al., 2018), research into performance-based training such as Evidence Based Training (EBT) (ICAO, 2013) and Horizon 2020's "FutureSky Safety: Human Performance Envelope" project (FutureSky Safety P6 Consortium, 2016).

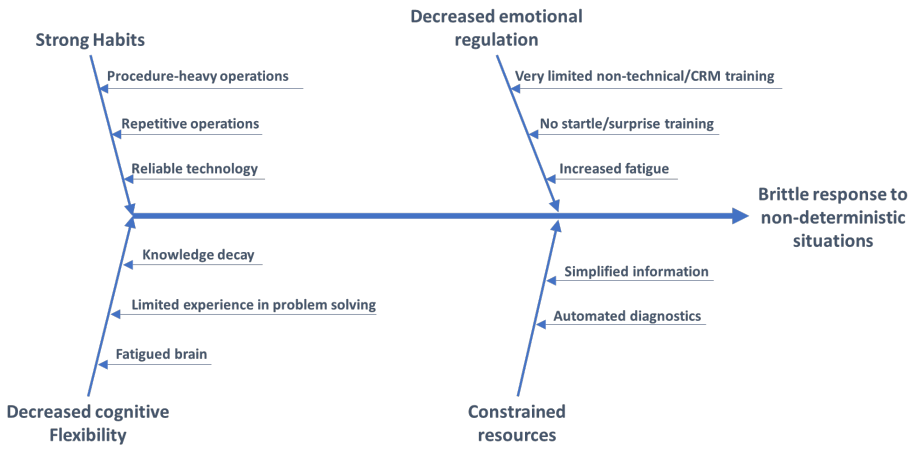


Figure 4.1. Contributory factors for brittleness in modern FDSTS response to complexity

Although this brittle FDSTS currently still maintains a high level of safety performance, problems arise when underlying design assumptions no longer hold true. Constant (1973) and Vincenti (1990) describe this as a *presumptive anomaly*:

> *"[A presumptive anomaly] occurs when scientific insight or assumptions derived from science indicate either that under some future conditions the conventional system will fail (or function badly) or that a radically different paradigm will do a much better job or will do something entirely novel." (Constant, 1973 – pg. 555)*

Research into pilot resilience, cognitive control and modern human-machine teaming showcase two types of anomalies that this FDSTS is presumed to be subjected to. The first (Type 1) presumptive anomaly is that the FDSTS will not maintain sufficient flight safety due to progressive, compliance-driven *erosion* of flight crew adaptiveness. Flight crew knowledge, awareness and flexibility degrade more than presumed due to prolonged exposure to high system reliability, operational consistency and procedural emphasis. This results in relatively high *experienced* opacity (even with known, but rare, system states). In analogy: the FDSTS has become more brittle than assumed.

The second (Type 2) presumptive anomaly is that this deterministic FDSTS will not maintain sufficient flight safety in modern and future aircraft systems featuring greater *true* opacity (due to increased complexity, integration and autonomy). This is because the FDSTS lacks the ability to *systematically* respond in a dynamic and unscripted way to non-deterministic situations. In analogy: future aircraft may require even greater "ductility" [16] than the current (brittle) FDSTS is designed for. Figure 4.2 depicts both presumptive anomalies in the complexity space. Figure 4.7 in Section 4.5 will illustrate seven notable accidents (actual anomalies) within this complexity space as early evidence of the presumed anomalies. By recognizing and appreciating these two anomaly mechanisms can a new—less brittle—JCS be considered as a new control paradigm.

[16] Ductility in the material sense: a physical property that measures a material's ability to undergo plastic deformation under stress without fracturing. I.e., changing shape in reaction to the specific stress applied.

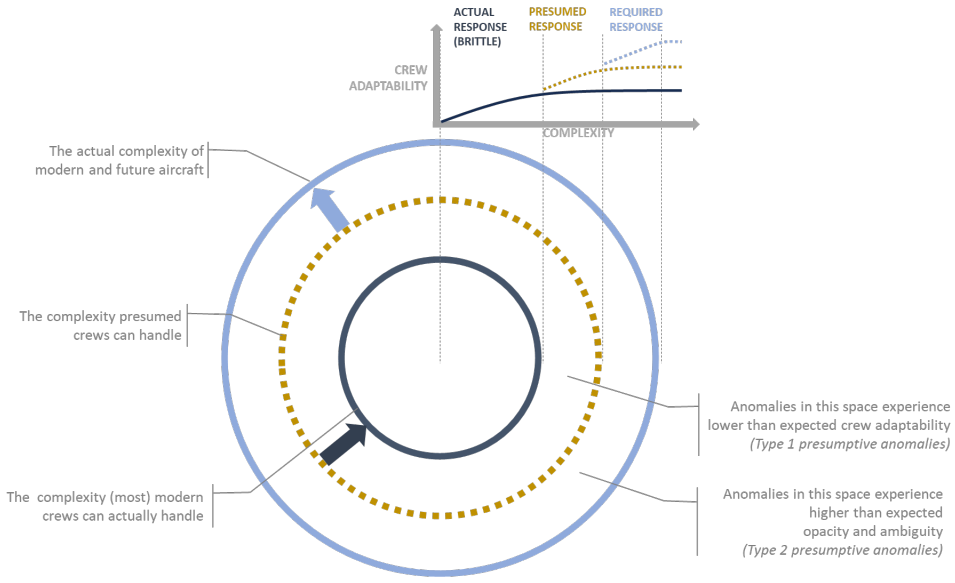


Figure 4.2. Visualization of FDSTS brittleness Type 1 and Type 2 presumptive anomalies

4.3 Solving for the brittleness problem

The two presumptive anomaly types described in subsection 4.2 share a common mitigation by means of reinforcing flight crews' ability to adapt their responses in opaque and ambiguous situations. This raises the question how to achieve this without disrupting the existing, compliance-based safety paradigm. The following subsections take a closer look at deterministic and non-deterministic control paradigms to identify how they may be assimilated.

4.3.1 The dichotomy of pilot control paradigms

In a general sense for all paradigms, pilots [17] have control when their *understanding*, *decisions* and *actions* reflect and resonate well with the specific nature of the situation they are experiencing. This is the basic concept of Neisser's Perceptual Control Cycle (PCM) (Neisser, 1976). In deterministic situations, past occurrences have resulted in checklists and procedures to streamline the process of recognizing a situation and selecting appropriate actions (e.g., memory items during a birdstrike on departure). The evolution toward such prescribed and trained recognition-primed decision making has been very effective for most (if not all) unambiguous operational situations (O'Hare, 1992; Li et al., 2014). Pilots proficient in such proceduralized problem-solving trust their training, apply procedures diligently and are

[17] The term pilot refers to the entire flight crew in a multi-crew JCS.

sensitive to recognizing and correcting for errors. Tony Kern's airmanship model (Kern, 2010), shown in Figure 4.3, provides an apt model of the pilot role in the deterministic control paradigm.

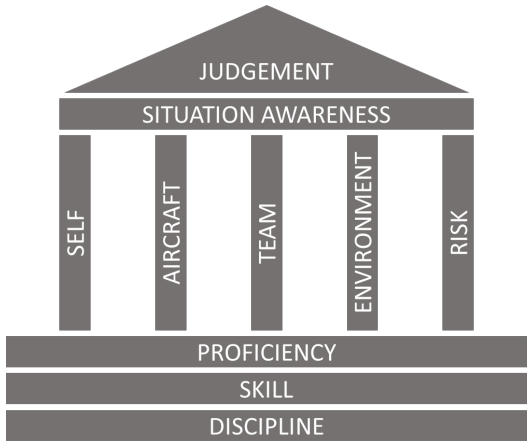


Figure 4.3. Tony Kern's Airmanship Model (Kern, 2010)

The apex of Kern's model is proper awareness and judgement, supported by five pillars of knowledge, which are in turn supported by a foundation of proficiency, skill and discipline. This model has a good fit for deterministic operations, where an upkeep of proficiency and knowledge supports an effective recognize-and-act way of working.

In non-deterministic situations, crews undergo the same basic process of understanding a situation, deciding on a course of action and executing accordingly, but without predetermined solutions. This goes beyond existing resolution frameworks such as T-DODAR, FORDEC and DESIDE (Banks et al., 2020), as these still emphasize "*recognize* the failure" and "performance *appropriate* actions" in situations which (by definition) lack known references and standards. Rather than a recognize-act response, the crew must cognitively engage in *sensemaking* (Hollnagel & Woods, 2005; Rankin et al., 2016).

Sensemaking is a generative process in which a person (or crew) assimilates raw information about the aircraft, flight path and situation into hypothesized scenarios that explain symptoms and observations. This is similar to the concept of "fluid intelligence" (Ziegler et al., 2012; Kent, 2017, Landman et al., 2017) described as the ability to arrange (new) variables into a coherent mental model. In contrast, "crystalline intelligence" is a stockpile of knowledge, insights and beliefs gained over a pilot's lifetime (like Kern's knowledge pillars). Where sensemaking relies strongly on fluid intelligence, crystalline intelligence (i.e., experience and knowledge) contributes to the sensemaking process through heuristic-based problem solving (Kahneman, 2011). Behaviors echoing sensemaking have

recently also been operationalized through the ICAO pilot competency framework (ICAO, 2013; EASA, 2020; IATA, 2023) in competencies such as "*Problem Solving and Decision Making*" and to a lesser extent "*Situation Awareness and Management of Information*".

Effective sensemaking requires higher cognitive functioning, which often decays quickly under stress (Field et al., 2018; Landman, 2020). For this reason, effective sensemaking requires a pilot to be aware of his/her own cognitive capacity and be able to re-engage it after a setback. Prime examples of this are startle and surprise recovery training (Field et al., 2018; Landman, 2020), increased emotional intelligence for collaboration (Matthews et al., 2003) and active fatigue prevention (Sallinen, 2020). In the ICAO pilot competency framework, "*Workload Management*" describes similar behaviors such as "*OB 8.9 Manages and recovers from interruptions, distractions, variations and failures effectively while performing tasks*" (ICAO, 2013). Table 4.2 illustrates the control dichotomy by contrasting the deterministic (prescriptive, compliant) paradigm versus the non-deterministic (sensemaking, learning) paradigm.

Table 4.2. Comparison of deterministic and non-deterministic control paradigms

	Deterministic control paradigm	**Non-deterministic control paradigm**
Safety by	Reproducibility	Adaptability
Most effective in	Transparent situations	Opaque situations
Promoted behaviors	Respect procedures	Challenge assumptions
	Remain within all limits	Focus on understanding
	Focus on execution	Generate options
Undesired behaviors	Non-compliance	Making assumptions
	Question procedures	No-cross checking
	Lack of punctuality	Lack of system interest/knowledge

4.3.2 The need for a singular pilot role

Both control paradigms have strong but different contributions to safety, although neither will suffice on their own. *Reproducibility* has been effective in reducing workload as increasingly demanding aircraft and flight operations strained pilot cognitive capacity. An overreliance on this paradigm, however, has resulted in the brittleness problem presented. *Adaptability* (sensemaking) provides a pathway in opaque and ambiguous situations. In turn, an overreliance on this paradigm would result in unreasonable workload as it is much more cognitively demanding than a procedural response.

Figure 4.4 illustrates how the ratio of these paradigms has approximately changed over time (blue), contrasted with the industry safety performance (green). The asymptotic limit seen in the safety performance (Maligne, 2011; Airbus, 2025) drives the re-appraisal of these control paradigms, particularly considering the two brittleness-related anomalies. Not doing so may even result in a decline in safety performance (purple) with more Type 2 anomalies. To ensure the safety performance improves beyond the current level (orange, upward trajectory), it is imperative that the new pilot role leverages both workload-reducing and adaptivity-enhancing mechanisms.

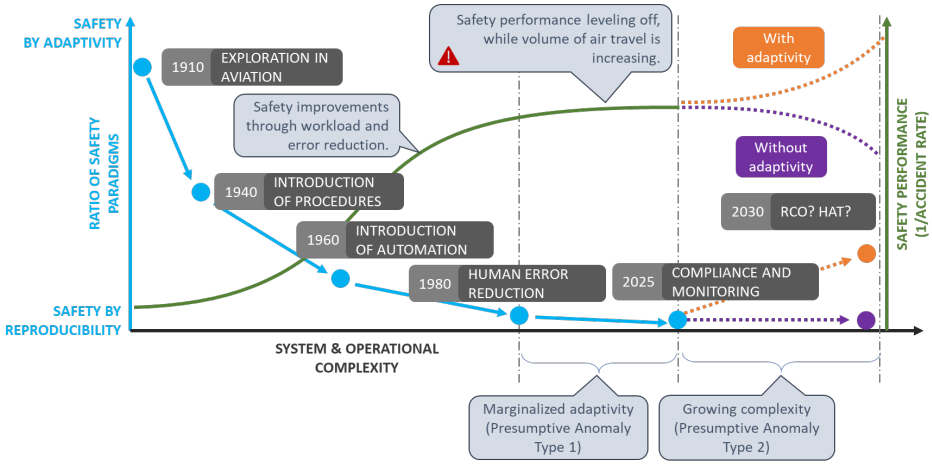


Figure 4.4. Conceptual visualization of control paradigm ratio (blue) and safety performance (green) over time with future scenarios (orange and purple)

There are three other arguments supporting a singular role rather than split paradigms. The first is the maintenance of recency and competence in both compliance and adaptiveness. As these paradigms are diametrically opposed, consistent use of one will result in the decay of the other (e.g., presumptive anomaly Type 1). A pilot role that consistently and repeatedly engages both compliance and adaptability ensures that pilots maintain proficiency in responding to both deterministic and non-deterministic situations. While the ICAO pilot competency framework (ICAO, 2013) appreciates behaviors from both paradigms, adaptive behaviors are a clear minority, most notably without any *systemic integration* to mitigate their decay. This "role-conflict" was already appreciated by Degani and Weiner (1994), describing it as an undesired "switching" between two operator modes.

The second argument is for removing standards duplicity for pilot behavior. The current strong emphasis on compliance presents accountability risks for pilots adapting and deviating from procedures, even when it may have improved the situation (Huijbrechts & Van Paassen, 2023). Unless it is established as part of the accepted behavioral framework,

the requisite adaptability behaviors discussed in Section 3.2 will be subject to a barrier as they are (often in hindsight) mutably labelled as a *merit* (e.g., in Quantas Flight 325F [18] and the DHL A300 missile strike in Baghdad (Maligne, 2004)) or an *error* (e.g., in Air Asia Flight 8501[6] and Asiana 214[6]).

The third argument for providing a single, unambiguous pilot role is its function as a design principle for efficient (re)design of the larger JCS. This allows for all human actors, systems and interactions to support the new pilot role rather than maintaining functionalities strongly engineered toward only one of the control paradigms. The DLR Risk Information System (RIS) developed in Man4Gen (Buch et al., 2017) or Reduced Crew Operations (RCO)-supporting automation as proposed by Ligda et al. (2015) are examples of such recent design efforts enabling a more dynamic pilot role.

4.4 Airmanship 2.0: A compound pilot role

The Airmanship 2.0 concept proposes an adaptation to Tony Kern's airmanship model by systematically *nesting* compliant behavior *within* a resilient attitude. This mitigates Type 2 presumptive anomalies (i.e., more frequent opaque operations) by structuring a *fail-safe* approach to the blend of transparent and opaque situations. In other words:

> *It is safer for flight crew [19] to assume opaque operations and, upon realizing that a situation is transparent, transition to a prescribed action, rather than assuming that an opaque situation is simpler than it is.*

This nesting is achieved through three extensions of Kern's model, as illustrated in Figure 4.5. The first is establishing a foundation of belief and *acknowledgement* of non-deterministic realities in modern flight operations. By voiding the illusion of a deterministic operation, pilots feature a wider safety aperture, which connects well with the principles of proactive, continuous Threat and Error Management (TEM) (Helmreich et al., 1999).

A second key adaptation is introducing the pilot's *cognitive self-regulation* to maximize cognitive facilities of oneself and others. This ability is essential to maintain high mental working capacity for sensemaking and engaging Kern's pillars of knowledge during opaque situations, and as such serves as a foundation for them in the model. It is placed above discipline, skill and proficiency as these capacities support cognitive regulation habits (e.g., meditation, breathing exercises, fatigue management). The increased propensity for startle

18 These cases are not examples of desired behavior but rather illustrate poorly scaffolded adaptive behavior. These cases will be discussed further in Section 4.5.

19 This same principle may also apply to other human actors facing potentially opaque situations in aviation and other domains.

and surprise that comes with highly reliable and complex systems (Martin et al., 2012; Field et al., 2018; Vlaskamp et al., 2025) calls for pilot cognitive recovery as an indispensable and explicit part of the new pilot role.

The third model adaptation requires *adaptability of mental models* as an essential foundation for situation awareness in opaque situations. This does not imply that pilots should doubt every mental model they have learned but that they recognize that these are only models with underlying assumptions, and that reality may prove to work differently. This attitude mitigates confirmation bias when seeking to re-establish situation awareness and invites a learning attitude. Such learning can be operationalized through practical sensemaking strategies as proposed in the Man4Gen project (Field et al., 2017; Mohrmann et al., 2017). Figure 4.5 illustrates these three adaptations to Kern's airmanship model.

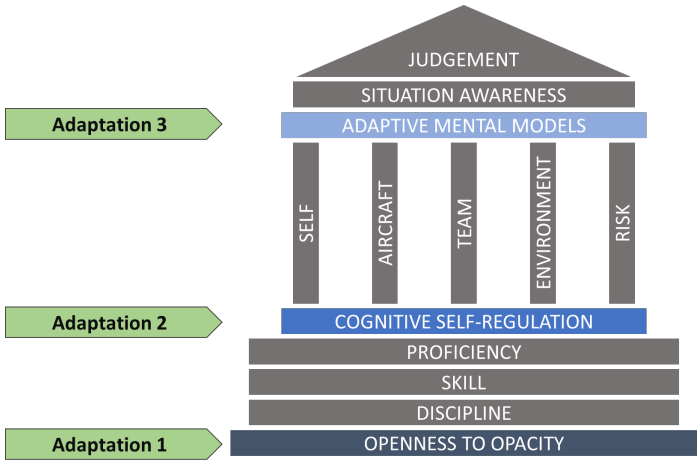


Figure 4.5. The Airmanship 2.0 adaptations of Kern's airmanship model

With this nested approach the Airmanship 2.0 concept aims to increase FDSTS adaptiveness while retaining established airmanship components. In keeping with the analogy to material brittleness and ductility: This nested approach is similar to compound materials featuring complementary performance through multiple substrates (e.g., rebar in concrete, or carbon fiber reinforced polymers), this concept facilitates the co-existence of contributions of both existing compliance (providing "strength") and new sensemaking behaviors (providing "ductility"). This can potentially result in three practical synergies between them. The first potential synergy is that the appreciation of opacity and mental model flexibility may invite aircrew to recognize critical (often implicit) assumptions underlying *diagnostic* procedures and actions, to improve their relative priority according to situational applicability.

The second potential synergy is that targeted deterministic responses (i.e., Standard Operating Procedures (SOP's), checklists, memory items) may improve the efficiency of the sensemaking process by reducing mental workload through manageable work items, shared situation/action awareness and reducing the number of slips and lapses. It is critical that such execution has an underlying *sensemaking objective*, and feeds back into the crew's situation awareness.

The third potential synergy may be that improved sensemaking of the current situation or system characteristics can help to identify which standardized *recovery* procedures and actions may be most appropriate, and which to avoid. Effective interventions may in turn continue to contribute to the (continuous) sensemaking process to reduce opacity. Figure 4.6 visualizes these three potential synergies between efficient (procedural) and sensemaking (adaptive) responses.

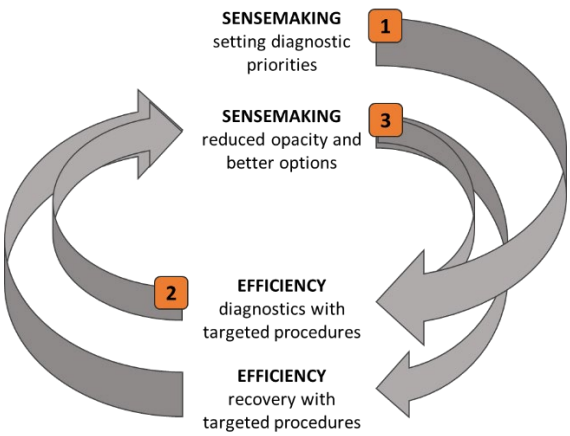


Figure 4.6. Airmanship 2.0 synergies between sensemaking and procedural efficiency

As a final note, it is worthwhile to consider the Airmanship 2.0 concept at the JCS level, particularly how systems can better support this new pilot role. For example: to maximize pilot cognitive capacity for sensemaking, high workload tasks including calculations and manual flight may be delegated to automation (ACROSS Consortium, 2016; Stanton et al., 2016). Aircraft systems could also proactively support mental model adaptation through varying the level of system details and providing trend data (Cahill et al., 2017). In the future, situation-sensitive systems (e.g., AI-based) may also support pilots in option generation and risk assessment (Kulida & Lebedev, 2020). Table 4.3 illustrates possible differences between the existing FDSTS and the Airmanship 2.0 JCS.

Table 4.3. High-level contrasting of the existing flight deck FDSTS with the proposed Airmanship 2.0 JCS

	Existing FDSTS	**Airmanship 2.0 JCS**
Safety by	Reproducibility	Adaptability & Reproducibility
Most effective in	Transparent situations	Opaque & transparent situations
Pilot role	▪ Monitor automation ▪ Manual flight ▪ Normal procedures ▪ Non-normal procedures ▪ Problem solving & decision making ▪ Optimize Flight	▪ Continuously build and improve understanding of the situation ▪ Identify and execute diagnostic queries ▪ Re-evaluate mission priorities ▪ (Re-)assess downstream options
Automation role	▪ Autoflight ▪ Warnings ▪ Status information (on demand) ▪ Autonomous system reconfiguration	▪ Autoflight ▪ Accelerate restructuring of mental models ▪ Support option-generation ▪ Transparent system reconfiguration

A review of the FDSTS will invariably invite a review of aircrew training programs, where existing competency-oriented, scenario-based training such as CBTA and EBT may serve a solid departure point for training preparing pilots for such an Airmanship 2.0 flight deck JCS. Section 5 discusses the Airmanship 2.0 concept by studying seven accident cases from the past two decades from the perspective of Airmanship 2.0.

4.5 Seven anomalies of opacity

Figure 4.7 reproduces Figure 4.2 with the placement of seven aviation accidents characteristic of Type 1 or Type 2 brittleness anomalies. There is one exception (Quantas Flight 32) which illustrates a so-called "black swan event" (Taleb, 2015), where the event likelihood is orders of magnitude smaller than current industry accident likelihoods.

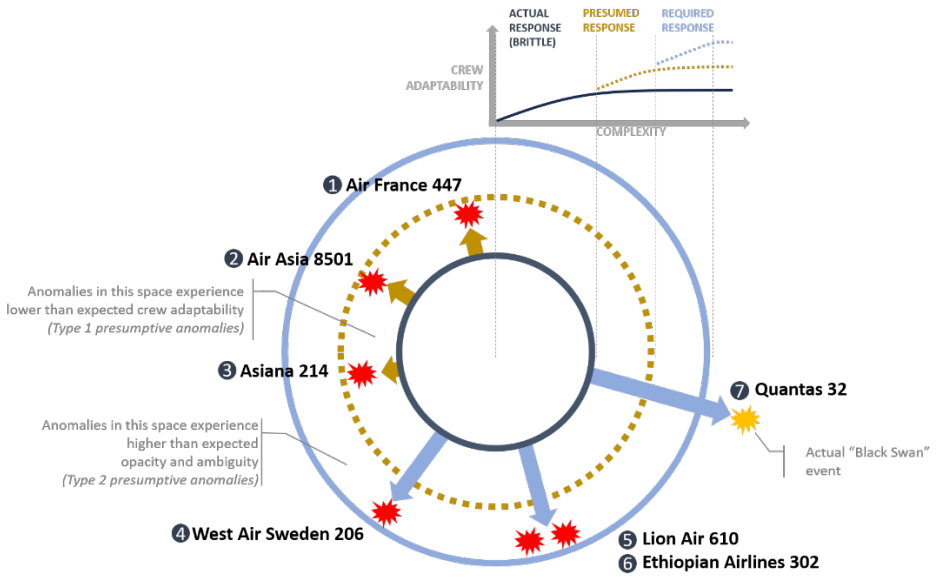


Figure 4.7. Illustration of seven accidents (anomalies) in the complexity space of Figure 4.2

Each case is first briefly summarized and then features a reflection on how the Airmanship 2.0 concept could have contributed to a less brittle response. The exception is Quantas 32 which illustrates a strong application of Airmanship 2.0 principles *avant la lettre*. The purpose of these cases is to revisit the underlying opacity and ambiguity flight crews experienced and how Airmanship 2.0 could have applied to such actual cases. The focus is not on the outcome and related crew actions and associated hindsight bias.

4.5.1 Air France 447

4.5.1.1 Synopsis (BEA, 2012)

An Airbus A330 experienced icing on the pitot tubes, resulting in inaccurate air data. The autopilot disconnected promptly, the aircraft control law switched from normal to alternate law, followed by the autothrust system disengaging. The pilots were unable to maintain control of the aircraft which was flying at an altitude where overspeed and stall speeds approach each other, combined with a removal of airspeed information. The resulting stall at night in low visibility was not detected, the aircraft crashed into the ocean.

4.5.1.2 Airmanship 2.0 reflection

Investigation into the crew functioning indicated a possible startle reaction (freeze) that may have caused inappropriate control inputs, possibly in reaction to greater *true* opacity (Type 2 anomaly). However, one year prior, the Airbus Operator Conference specifically addressed 1) the issue of a loss of airspeed indication, 2) the startle reaction that it could

cause, providing guidance to operators for loss of airspeed indication in both operations bulletins and training guides (BEA, 2012). As such, this accident may have instead featured greater *experienced* opacity without higher true opacity (Type 1 anomaly). Despite this, the removal of a primary flight metric (airspeed) introduced ambiguity about the aircraft's attitude, speed margins (with stall speed and overspeed limits in close proximity at such cruise altitudes) and autoflight capabilities hours aways from the nearest airfield. The disappearance of such a primary indicator may well induce surprise and possibly even startle, requiring pilots' cognitive and emotional recovery before engaging in sensemaking and taking further action. Such self-restoration may be counter-intuitive when warning systems and disengaging autoflight systems are indeed "panicking" without airspeed indications, but a recovery allows crews to re-engage in bigger picture sensemaking (energy management and energy margins) and prevent over-focus on (narrow-scoped) system warnings. Several BEA report recommendations (BEA, 2012) align with Airmanship 2.0, notably training of surprise management skills and improving the design of flight deck information and warning systems to reduce the opacity of a situation and provide crews with more, clearer cues to diagnose the situation, even for such extremely rare, *but known*, situations.

4.5.2 Air Asia 8501

4.5.2.1 Synopsis (KNKT, 2015)

An Airbus A320 experienced a malfunction of the Rudder Traveler Limiter Unit (RTLU), later attributed to a tiny crack in a soldered connection. This specific aircraft experienced a similar RTLU malfunction 23 times in the previous year. Airbus procedures require resetting the RTLU. During this flight, the system presented an RTLU failure four times. After the fourth time taking prescribe ECAM reset actions, the captain elected to reset the two Flight Augmentation Computers (FAC) using the circuit breaker, placing the aircraft in alternate law and manual control. Shortly after, a small, sustained rudder deflection resulted in an undetected roll, after which the aircraft entered a stall and crashed.

4.5.2.2 Airmanship 2.0 reflection

Similar to the Air France 447 accident, the chain of events begins with a small component failure (solder crack) causing a great deal of opacity in the flight deck, showcasing the highly leveraged nature of complex systems. With ECAM actions not effective in resolving the recurring RTLU fault, the crew was presented with both opacity ("what is going on in the system?") and ambiguity ("reset to recover rudder or leave it and continue with limited rudder control?"), particularly with the deteriorating weather in the area. While the report notes contributory factors such as insufficient Upset Prevention and Recovery Training (UPRT), poor crew coordination (e.g., no pilot flying/monitoring role divisions) and poor Threat and Error Management (TEM) briefings, it was also suspected that the

captain's non-standard FAC-reset was based on a previous maintenance action to reset the RTLU with a circuit breaker reset and confirming that this could be done every time the failure was presented on the ECAM (KNKT, 2015). While this may have been permitted on the ground, Airbus did not list this reset as permitted during flight, stating that such a reset must be carefully considered. While seeking other resolutions when prescribed actions are ineffective can be regarded as resilient or adaptive behavior when prescribed actions are ineffective, it must be sufficiently scaffolded in a sensemaking process that considers threats and contingencies. An Airmanship 2.0 approach would activate a sensemaking process at the very first ECAM warning indicating an RTLU failure, where prescribed ECAM actions would be part of the sensemaking process. By seeing actions as a way to make sense of the situation, rather than only a resolution activity, may have elicited a more thorough appreciation of the threats and risks of a non-standard circuit breaker reset before engaging in it. This may have reduced the crew's strong *switching* from compliance to "creativity" without validating their adapted mental models before acting. Furthermore, an improved appraisal of the situation by evaluating flight continuation options (e.g., could the flight proceed to the destination or diversion airfield with an unresolved RTLU fault?) may have identified other means to maintain safe flight rather than focusing on RTLU resets. A JCS improvement similar to the Man4Gen RIS (Buch et al., 2017), discussed in subsection 3.2, could help identify the *actual* controllability risks.

4.5.3 Asiana 214

4.5.3.1 Synopsis (NTSB, 2014)

A Boeing 777 on final approach descended below the glideslope and impacted the sea wall just before the runway threshold, resulting in three fatalities. During the descent air traffic control requested the crew to maintain a higher speed until closer to the airport. As a result the aircraft few above the glideslope. The crew made successive changes to the aircraft's vertical speed autoflight modes to return to the glideslope, resulting in the use of Flight Level Change (FLCH) mode which is not part of Asiana's approach SOPs. This mode, in combination with a thrust lever manual override to idle thrust, disengaged the autothrust system, putting it into an autothrust "hold" mode without low-speed protection modes. As the aircraft descended toward the glideslope from above, the crew let it recapture the glidepath at around 500ft AMSL, after which the aircraft slowly bled speed as the autothrust was still in a hold mode, resulting in a low energy state which was recognized too late for a successful go-around.

4.5.3.2 Airmanship 2.0 reflection

The sequence of events and automation responses was clearly traceable without any indication of a system failure. However, the fact that the crew's non-standard use of autoflight and autothrust modes resulted in a situation they themselves could not manage, despite the aircraft being fully serviceable in good weather conditions, illustrates the brittleness that may arise through current crew conditioning (Type 1 anomaly). The investigation also underscores the complexity of current autoflight and autothrust systems in relation to the level of details flight crews are trained on. The foundation of the Airmanship 2.0 *openness to opacity* applies well to this situation, where a general appreciation that the aircraft may feature more complexity than the flight crew's own familiarity with it. Such an approach to the aircraft would be able to elicit a stronger threat evaluation during the approach which could in turn drive effective mitigation such as enhanced autoflight mode monitoring, preparing for a missed approach decision or requesting a more familiar approach profile from air traffic control. This case highlights the pro-active, attitudinal aspect of Airmanship 2.0, rather than its sensemaking aspects highlighted by other cases.

4.5.4 West Air Sweden 206

4.5.4.1 Synopsis (SAIA, 2016)

During a cruise flight at night a Canadair CRJ200 with the autoflight system engaged, the left-seat Primary Flight Display (PFD) showed a sudden, very rapid increase in pitch attitude due to a failure of one of the attitude reference systems. This automatically disengaged the autopilot (which is coupled to the left-seat instruments), requiring the crew to take immediate manual control. The PFD initially provided a very brief indication of a "*pitch angle disagreement"* between the left and right PFD. However, as the indicated pitch angle rapidly increased over a certain limit, a PFD *declutter mode* activated and removed this pitch disagreement indication, also adding a "push down" indication on the PFD. The captain obliged immediately, resulting in the aircraft entering an uncontrolled dive and impacting the ground less than two minutes later.

4.5.4.2 Airmanship 2.0 reflection

This case showcases how system integration complexities can result in undesired, emergent system behavior. The pitch-induced disconnection, pitch-disagreement indication and declutter functions were all each designed from a safety perspective. However, their interaction likely presented the crew with a particularly opaque situation, exacerbated by a call for immediate action (i.e., a "push down" indication). While similar to the AF447 case in terms of information degradation and reduced autoflight capabilities, this Type 2 case featured greater *true* opacity, showcasing the need for crews to be prepared for (safety) systems to act in unpredictable ways (openness to opacity). However, the similarly with

AF447 also lies in the need for flight crew to be aware of the need for personal recovery in light of acute and unexpected developments. Such a capacity would have dampened crew actions to allow for a brief window of sensemaking (e.g., for visual and vestibular sensing) prior to taking action. From a JCS perspective, more graceful system degradations (Woods, 2015) could also cater for a safe transition from automated control to human control and sensemaking.

4.5.5 Lion Air 610 & Ethiopian Airlines 302

4.5.5.1 Synopsis (KNKT, 2019; EAIB, 2022)

Two Boeing 737MAX aircraft exhibited very similar failures of its Maneuvering Characteristics Augmentation System (MCAS). Both investigations suspected malfunctions of the single Angle of Attack (AOA) sensor inputting to the MCAS which subsequently trimmed the aircraft nose down. Crew attempts to pull the nose up resulted in multiple, cumulative MCAS nose-down inputs, until the aircraft pitch was unrecoverable. In both cases the crew was simultaneously confronted with airspeed- and altitude-disagree messages, master caution warnings and a stick shaker response, possibly inducing a startle response. Both crews were unable to correctly execute the non-normal checklist to disengage the pitch trim system. Neither aircraft was equipped with an optional *angle-of-attack disagree* warning light. A few days before the Lion Air Flight 610 accident, Lion Air flight 43 operated the same aircraft as Lion Air 610 and also experiencing the same MCAS pitch down and alerts (KNKT, 2019). This crew was able to operate the pitch trim disconnect switches in time. This recovery took minutes with significant pitch control challenges, rather than several seconds as was assumed in the 737MAX certification program (Senate Committee, 2020). However, as Flight 43 was only an incident it unfortunately was not subject to a full investigative report as the following accident flight.

4.5.5.2 Airmanship 2.0 reflection

Like the West Air Sweden 206 case, the MCAS system likely contributed to system opacity well beyond the competence of today's aircrew (Type 2 anomaly). In particular, the repeated and cumulative nose-down pitch trimming resulted in a near-reversal of longitudinal control: pulling to raise the nose only led a stronger force pushing it down. The assumption that crews would recover from an MCAS failure by identifying it as a stabilizer runaway doesn't hold as such a runaway provides a loud, continuous trim wheel activation (strong cue), while the MCAS pitch changes were much more subtle and progressive. Lastly, the lack of pilot training, system knowledge and crew notifications likely shrouded a critical control malfunction (true opacity). Both accident cases suffered further from uncoordinated crew responses, likely startle induced and possibly even fear-potentiated (Field et al., 2018) due to progressive nose down trimming. The Lion Air crew did not communicate about the control

difficulties they were struggling with (possibly a freeze response), and the Ethiopian Airlines crew repeatedly attempted to engage the autoflight system rather than responding to the warnings and alerts they were experiencing (possibly a fixed mental model). The Airmanship 2.0 concept may have improved outcomes through prioritizing (some) cognitive recovery and establishing some action feedback as part of a sensemaking process, rather than repeating the same actions (e.g., trying to reactivate the autoflight system) and hoping for a different outcome. A more salient report about the preceding Lion Air Flight 43 incident may have provided more insight into the crew's experiences, mental models and resolution of MCAS opacity. From a JCS perspective, greater transparency in systems goes a long way in helping flight crews differentiate between multiple situations with similar symptoms. Since these accidents, the AOA disagree warning indication has become standard rather than optional (Boeing, 2019), which is a small step in this direction.

4.5.6 Quantas 32

4.5.6.1 Synopsis (ATSB, 2013)

An Airbus A380 experienced an uncontained engine failure in cruise flight. A separated fan blade damaged multiple hydraulic, electric and structural systems in the left wing, causing malfunctions in flight-, engine- and fuel control systems with the ECAM displaying 53 faults. In comparison, pilot training scenarios deemed challenging by proficient crews may feature only two or three ECAM faults. By coincidence, this flight featured three additional experienced crew members performing trainer qualification checks. After two and a half hours in the vicinity of Singapore diagnosing the state of the aircraft, the crew managed to land successfully with one engine still uncontrollable near max thrust, reduced hydraulic braking, failed antiskid and only thirty minutes until a lateral fuel imbalance would have resulted in complete loss of control.

4.5.6.2 Airmanship 2.0 reflection

This case qualifies as a true "black swan" (Taleb, 2015) event as it was considerably less probable than current certification standards for catastrophic events (10^{-9}) as per EASA AMC CS25.1309-7 (EASA, 2023). According to Captain De Crespigny the likelihood was nearer to 10^{-14} (De Crespigny, 2012). Unfortunately, the ATSB report is very thin on human performance, merely stating the crew worked as a team and followed SOPs (ATSB, 2013). De Crespigny's own account (De Crespigny, 2012) features much more detail about the crew's cognitive processes and experiences. Despite the considerable opacity and ambiguity experienced by the crew, the crew's ability to diagnose system states, create an *adapted* mental model of the aircraft and identify a novel recovery plan is a strong example of Airmanship 2.0. Their use of ECAM and other procedures as *part* of their sensemaking process drove efficient and effective adaptation of their mental model of the aircraft's capabilities.

Ultimately, reducing the opacity by clearly establishing *what functionality they still had* (rather than focusing on what failed) most likely gave them a much more actionable sense of the situation. This is a great example of an adaptive mental model, which is reconstructed bottom up, rather than maintaining mental models of systems and the aircraft which are not accurate in that context. A final key element is effective time management to facilitate sensemaking. In line with recommendations and results from Mohrmann et al. (2015) and Mohrmann (2026), managing short term planning to create time (e.g., by entering a holding pattern) provides both temporal and cognitive space for sensemaking, reconfiguring and mitigating downstream options and threats. In effect, time management is a precursor to most of the Airmanship 2.0 principles: allowing time to recover cognitively, allowing time to think/reflect and allowing time to perform actions and reflect on their impact. Other accidents such as Lion Air 610 and West Air Sweden 206 also provide similar contexts where building in time may have resulted in better outcomes.

It must be mentioned that the above accidents may have been (partially) mitigated with other non-aircrew countermeasures such as improved safety management (e.g., preceding instances of MCAS and frozen pitot tubes) or automation interaction (e.g., crew notifications). However, such hindsight does not generate sufficient technical foresight to void the requirement for resilient flight crews. Opacity and ambiguity are already a reality and will increasingly be part and parcel of modern flight operations. A reappraisal of the pilot role is essential to mitigate what cannot always be anticipated (without an accident to point it out).

4.6 Discussion

This chapter revisits basic premises of human-automation teaming in complex, safety-critical systems to reduce systemic brittleness. It brings to the surface critical underlying assumptions and non-sequiturs in introducing adaptive pilot responses in a currently compliance-centered paradigm. The Airmanship 2.0 concept is proposed, which leverages the particularly human capacity of sensemaking to adapt in the face of opacity and ambiguity. It systematically retains existing compliance safety performance by nesting existing workload-reducing behaviors (e.g., procedures) within the process of sensemaking. By defining a singular, compound role, the concept aims to (1) prevent pilot skill decay in both compliance and adaptivity, and (2) avoid straining pilots in role-bifurcation. Successful introduction of this concept will likely require a more succinct analysis and appreciation of the various JCS interactions and automation roles potentially supporting a pilot model based on Airmanship 2.0, as well as the potential impact on flight operations and aircrew training.

The Airmanship 2.0 concept is relevant for today's Type 1 and Type 2 brittleness anomalies, but even more so for two significant developments on the horizon. The first is

Reduced Crew Operations (RCO) exploring a single-pilot JCS (Sprengart et al., 2018; Schmid & Stanton, 2020) as well as concepts with ground-based pilots (Niedermeier et al., 2023). These concepts explore shifting more tasks to automated systems (Shively et al., 2016; Ho et al., 2017), further propagating Klein's self-reinforcing complexity loop and potentially increasing the risk of Type 1 brittleness anomalies. Harris (2023) also calls for the reappraisal of the pilot role for RCO concepts, shifting from executive activities to strategic, decision-making activities. RCO potentially also increases the likelihood of Type 2 anomalies as the effective cognitive capacity for sensemaking in the flight deck has reduced by (at least) half. Furthermore, RCO potentially affects the sensemaking response by removing valuable human interaction during 1) cognitive recovery (i.e., startle management crew cross-checking as described by Field et al. (2018)), 2) the mutual challenging of personal biases and 3) mitigation of (cognitive) blind spots (Sarter et al., 2023). This is not to say that RCO should not be explored, but it does underscore the need to explicitly design required crew behaviors (including adaptive responses) into a new (Airmanship 2.0) JCS, rather than inheriting the modus operandi from existing multi-crew flight decks.

The second development on the horizon is the introduction of AI-based systems in aviation (EASA, 2023; FAA, 2024). The potential for significant autonomy of automated systems (Kirwan, 2025) brings with it considerable risk of opacity and complexity and increased risk of Type 2 anomalies. This in part is driving robust and paced integration as proposed in EASA and FAA AI roadmaps. Research into human-AI teaming underscores limitations of AI to fully replicate the human ability of causal mental modelling (Sarter et al., 2023) through cognitive processes such as critical thinking, assumption testing, and intuition. As AI agents enter the flight deck JCS, a reappraisal of the pilot role toward such cognitive capacities is well called for. This may be particularly relevant in combination with RCO concepts (Minaskan et al., 2022). In contrast, the Airmanship 2.0 concept may also serve as an operator model for other safety-critical systems such as Air Traffic Management (ATM), which faces the integration of unmanned aviation and autonomous non-piloted systems together with increased levels of automation, possibly requiring a reappraisal of the human role in traffic management (Malakis et al., 2023; Zou, 2025).

As a final reflection: explicating the brittleness problem, its presumptive anomalies, a revised pilot role and future developments around RCO and AI raises the question: how can the industry identify in time when Airmanship 2.0 also reaches its limits? When would we need Airmanship *3.0*? This would be the hallmark of *industry-level resilience*. Kirwan (2025) provides a clear histological overview showing how human factors methods have evolved as the flight deck's socio-technical system evolved. Traditional models such as Human Factors Analysis and Classification System (HFACS) don't provide sufficient resolution to address the cognitive and behavioral nuances that drive today's flight deck performance. More

fundamentally, the research field of "human" factors may be too narrowly scoped for these nuances, possibly requiring an evolution toward a systems perspective at the JCS level as suggested by Woods (2019).

One attempt to proactively defend against future blind spots is the System Theoretic Accident Model and Processes (STAMP) approach, and in particular its applications to design with Systems-Theoretic Process Analysis (STPA), and to accident investigation with Causal Analysis Using System Theory (CAST) (Leveson, 2016). By mapping entire systems "as they are", the STAMP approach attempts to uncover interactions, behaviors and sensitivities not visible to frameworks scoped around the human. In an industry with such a high level of safety, reactive incident- and accident-led learning may not provide sufficient evidence to understand modern FDSTS criticalities and sensitivities in time. Rather, the expansion of scenario-based training through CBTA (ICAO, 2013; IATA 2024) may provide the requisite immersion and operational relevance to better understand these cognitive nuances in the FDSTS. As a first step in this direction, the IATA white paper on EBT already proposed that the pilot competency framework replace the existing flight crew countermeasure safety taxonomies (IATA, 2024a). As the near future herald's radical operational concepts for tomorrow's flight decks, the industry must be more vigilant than ever on the continued validity of design assumptions inherited from the past.

5

Making sense of sensemaking

Positioning sensemaking as the default mode of engagement through the Airmanship 2.0 model invites a deeper understanding of the neurocognition of sensemaking in order to identify trainable processes and behaviors, as well as understand possibly operationalization aspects. This chapter provides a breakdown of the neurocognitive processes responsible for sensemaking in various modalities of opaque operations. It presents an initial simulator study to evaluate how such neurocognitive training may improve flight safety through reinforcing key competencies related to sensemaking. The implication of such training is discussed.

The content of this chapter is based on:

Paper title	Why we need more showers in the aircraft: A neurocognitive perspective on the future pilot
Author(s)	Mohrmann, J.F.W.
Published in	Pre-print (engrXiv, 2026)
DOI	https://doi.org/10.31224/7073

ABSTRACT *High levels of complexity in modern commercial aircraft and flight operations feature new accident types featuring a brittle flight crew response to increasingly frequent ambiguous and opaque non-normal situations. The previous chapter from the author proposes the Airmanship 2.0 concept to mitigate this brittleness by harmonizing the existing compliance-centered safety paradigm with an adaptive, sensemaking function in the flight deck. This chapter explores the neurocognitive sensemaking mechanisms that may enable this compound role, including related training considerations. A case study in mindfulness training is discussed as an initial practical exploration into such neurocognitive training of pilots. There are tentative indications of a positive impact of mindfulness on flight crew response to ambiguous situations, alongside key reflections on aviation cultural biases and (lack of) readiness for such advanced training.*

5.1 Introduction

The aviation industry's safety performance is very high. However, in the past two decades it has been struggling to continue reducing the accident rate (Airbus, 2025). Both accident investigations and safety research are identifying a new accident modality related to undesired interactions between increasingly complex aircraft systems and a strong emphasis on compliance (Mosier et al. 1998; Manzey et al., 2012; IATA, 2014b; Man4Gen Consortium, 2015; Field et al., 2018; Woltjer et al., 2015; Banks et al., 2020; Landman et al., 2020; Hancock et al, 2022; Sarter, 2024; Mohrmann, 2026). Chapters 2, 3 and 4 have explored the nature of these accidents and identified key differences between crews that effectively navigate such complexity and crews that struggle to maintain safe flight operations in this context. Differences are mainly found in crews' abilities to make sense of an ambiguous situation and adapt beyond a purely procedural response.

Such recovery behaviors contrast with the current problem-solving behavioral standards for flight crew which are centered on monitoring and procedural compliance. The Airmanship 2.0 concept described in Chapter 4 addresses this bifurcation through a compound pilot role, but in turn elicits a deeper question about the actual neurocognitive basis to interweave both compliant and adaptive behaviors. In other words, how may the concept of Airmanship 2.0 actually manifest in the cognition, neurology, cognition and biology of the human pilot?

This chapter unpacks the neurocognitive processes related to sensemaking of ambiguous situations, considerations in training and developing such processes and presents a practical case study in neurocognitive training of pilots for ambiguous and complex flight operations. Section 5.2 revisits the sensemaking challenges and Airmanship 2.0 concept as a

point of departure, Section 5.3 explores the underlying neurocognitive mechanisms of sensemaking and Section 5.4 describes how these may be trained and applied in the context of flight operations. Section 5.5 presents a case study exploring such neurocognitive training in pilots, followed by a discussion of research, design and operational considerations associated with the proposed pilot control paradigm.

5.2 Safety in complexity

This section revisits the transitional challenge related to the pilot role and the proposal of a sustainable and harmonized *sensemaking function* through the Airmanship 2.0 concept. The reader is referred to Chapter 4 for a more succinct examination of the theoretical and casuistic foundation of the Airmanship 2.0 concept.

5.2.1 A bifurcated pilot role

The very first pilots in the early days of initial fixed wing flight were avid cyclists: their functional role was to power the aircraft and manage basic lateral and vertical flight controls (Chant, 2002). In over a century since, the pilot role has evolved considerably beyond cycling, with the introduction of checklists, procedures, multi-crew operations and automation: all to accommodate the ever-increasing performance and efficiency requirements of a growing and developing industry. However, these developments also increased the complexity of systems and operations that pilots are managing to such an extent that pilots are (increasingly) detached from a deeper familiarity with and knowledge of aircraft systems (Woods, 2004; Langewiesche, 2009; Man4Gen Consortium, 2012; Saurin & Carim, 2012; IATA, 2014b; Man4Gen Consortium, 2015; Stoop & Van Kleef, 2015; Woltjer et al., 2015; Prinzel et al., 2024). This introduced new human performance issues such as automation bias, monitoring errors and knowledge degradation (Weiner & Curry, 1980; Bainbridge, 1983; Sarter & Woods, 1994; Parasuraman & Riley, 1997; Mosier et al., 1998; Manzey et al., 2012; Mohrmann et al., 2015; Strauch, 2017; Kharoufah et al., 2018; Proctor & Van Zandt, 2018; Kelly & Efthymiou, 2019; Gago et al., 2025). To mitigate such "pilot errors", the industry furthered the shifting of tasks from pilots to automation as well as reinforce crew compliance to procedures, driving for safety through reproducibility.

While such reproducibility achieves significant safety performance, this safety paradigm is seemingly at its limits. In light of this, flight safety research (IATA, 2014b; Man4Gen Consortium, 2015; Woltjer et al., 2015; Field et al., 2018; Banks et al., 2020; Landman et al., 2020; Hancock et al., 2022; Sarter, 2024; Mohrmann & Field, 2026) in the past two decades has been pointing to a relative increase in (crew-experienced) ambiguity and opacity related to the complexity and reliability of modern (e.g., fourth generation) aircraft, coupled with compliance-conditioned flight crews struggling to make sense of and

resolve such situations which are unfamiliar and non-deterministic. The growing complexity of aircraft systems has, ironically, resulted in flight crew not proficient in managing that complexity. The desired resilient response to complex, ambiguous and opaque situations enables creative, dynamic problem-solving which can be described as the process of *sensemaking* to formulate effective responses to an opaque situation (Hollnagel & Woods, 2005; Mohrmann et al., 2015; Rankin et al., 2016; Mohrmann et al., 2017; Sarter, 2024).

This introduces a bifurcation in the pilot role: on the one hand requiring a high level of procedural compliance and basic task proficiency (e.g., lowering the gear), while on the other hand requiring pilots to demonstrate acute, dynamic sensemaking and problem solving in rare (but often critical) opaque and ambiguous situations lacking prescriptive solutions. This is often sardonically referred to by pilots as *"…endless hours of sheer boredom, punctuated by moments of stark terror."* (Diehl, 2013). The near opposition of these control paradigms features contra-reinforcement (e.g., heavy use of compliance decays adaptability), mutable behavioral standards and unidirectional optimization of Joint Cognitive System (JCS) designs (Mohrmann, 2026). As such, these phenomena void the (implicit) assumption that pilots are readily proficient in both control paradigms and call for explicit harmonization of both paradigms into a single, non-bifurcated pilot role. Mohrmann & Stoop (2019) proposes the Airmanship 2.0 concept to achieve this.

5.2.2 The Airmanship 2.0 concept

The Airmanship 2.0 concept proposes an adaptation of Tony Kern's airmanship model (Kern, 2010). The apex of Kern's model is proper awareness and judgement, supported by five pillars of knowledge, which are in turn supported by a foundation of proficiency, skill and discipline. As such, this model aptly describes the pilot role for transparent operations, where an upkeep of proficiency and knowledge supports an effective recognize-and-act way of working, such as Klein's (1996) Recognition-Primed Decision-making (RPD) model. The Airmanship 2.0 concept adapts this model by systematically *nesting* compliant behavior *within* a sensemaking approach. This follows a *fail-safe* approach to the blend of transparent and opaque situations crews may encounter. In other words:

> *It is safer to assume opaque operations and, upon realizing that a situation is transparent, transition to a prescribed action, rather than assuming that an opaque situation is simpler than it is.*

The Airmanship 2.0 concept leverages innate human competence in sensemaking to learn and adapt in the face of opacity and ambiguity. Critically, the proposed concept retains existing compliance safety performance by nesting existing workload- and error-reducing behaviors (e.g. procedures) within this process of sensemaking. By defining a unified yet compound role, this concept prevents pilot skill decay by design in both

compliance and adaptivity, while avoiding straining pilots in role-bifurcation. The proposed nesting is achieved through three adaptions to Kern's model, as illustrated in n Figure 5.1.

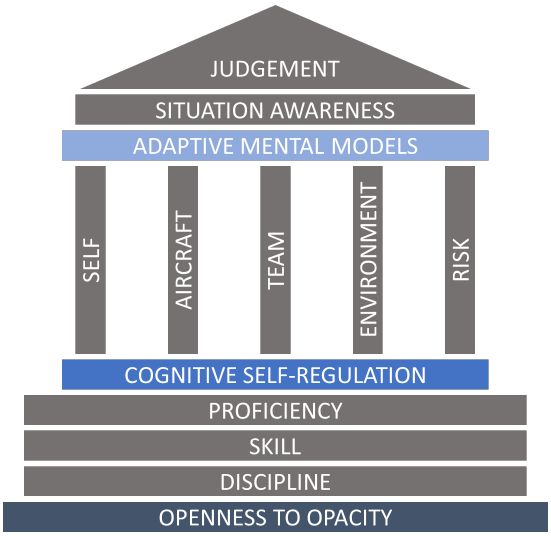


Figure 5.1 The Airmanship 2.0 concept (Mohrmann, 2026)

The first is establishing a foundation of belief and *acknowledgement* of non-deterministic realities to widen a pilot's safety aperture and support proactive, continuous Threat and Error Management (TEM) (Dekker & Lundstrom, 2006). The second adaption is the pilot's *cognitive self-regulation* to maximize cognitive resources required for sensemaking and (the pillars of) knowledge during opaque situations. This adaption also critically mitigates the increased propensity for startle and surprise that comes with highly reliable and complex systems (Martin & Sloman, 2013; Rivera et al., 2014; Field et al., 2018; Landman et al., 2020; Gardini et al., 2022; Vlaskamp et al., 2025).

The third adaption is maintaining *adaptivity in mental models* to make sense of non-deterministic situations. This does not imply that pilots should doubt every mental model they have learned, but that they recognize that these models are limited to underlying assumptions, and that reality may prove to work differently. This attitude mitigates confirmation bias and could be operationalized through practical sensemaking strategies as proposed in the Man4Gen project (Field et al., 2017; Mohrmann et al., 2017). Readers are referred to Chapter 4 for further details and reflections of the Airmanship 2.0 concept.

These three proposed adaptions to the airmanship concept are believed to facilitate a sensemaking response that is symbiotic to the deterministic response (i.e. compliance-based response). While this outlines desired flight crew response at a conceptual level, it fails to articulate how this manifests cognitively, and – critically – how this in turn can be conditioned

and trained. This requires a deeper appreciation of the human neurocognition of sensemaking (particularly under ambiguous circumstances), which is the subject of Section 5.3.

5.3 The Neurocognition of sensemaking

The neurocognition around learning sand specifically sensemaking is complex to say the least, and a full and comprehensive review of the science of cognition is beyond the scope of this work. This section will explore the main neurocognitive processes around sensemaking, with a particular focus on the context of ambiguous and complex situations. Section 5.3.2 will reflect how these processes facilitate the compound nesting proposed in the Airmanship 2.0 concept.

5.3.1 Different neurocognitive pathways for sensemaking

At a basic level, our brain makes sense of the world through a perpetual cycle of processing inputs that modify an existing understanding or model of the world, which in turn directs our decisions and actions, a basic process described by Neisser (1976). However, at a neurological level, this basic concept of cyclic sensemaking can run through very different neural networks and brain cortices and have their independent merits and drawbacks. As a point of departure, it is worth appreciating three long-held and widely studied theoretical constructs of our cognition and sensemaking mechanisms:

1. Skill-rule-knowledge based behavior (SRK) (Rasmussen, 1983)
2. Recognition-primed decision making (RPD) (Klein, 1996)
3. Dual process theory (DPT) (Kahneman, 1982)

Table 5.1 provides a high-level summary of these constructs. All three constructs have in common a basic appraisal of both autonomous, intuitive sensemaking as well as conscious, conceptual, analytical sensemaking, a bifurcation aptly described by the French philosopher Blaise Pascal as "*esprit de finesse"* (intuitive mind) and "*esprit de geometrie"* (analytical mind) (Pascal, 1670/1995; Barrouillet, 2011). Furthermore, they all underscore the close coupling of sensemaking and decision making: our decisions are indeed based on our (possibly inaccurate) understanding of the situation. This is important as sometimes the decision-making context drives a particular sensemaking strategy (and vice versa), which will be discussed below.

Table 5.1. Summary of three legacy theoretical sensemaking constructs

Construct	Description
SRK Construct	Behavior occurs at three levels: *skills* (sensory-motor/reflex response), *rules* (pre-determined responses) and *knowledge* (analytical response).
RPD Construct	Understanding ("frame") of the situation is initially mapped heuristically to previous experiences. Mapping errors trigger an analytical/reframing response to create a new frame.
DPT Construct	The brain operates by default on a fast and frugal mode that uses heuristic thinking (System 1) but can also engage a cognitively more demanding analytical response (System 2) if System 1 delivers errors.

Behavioral and neurological research in the past decades provides a more succinct appreciation of our cognitive sensemaking and decision-making processes. Five key nuances to these three legacy models of cognition (relevant for the context of modern aviation) are articulated in the following subsections.

5.3.1.1 Nuance 1: Multi-faceting of "intuition"

The sensemaking that happens outside of our conscious experience seems to be more complex than simply a fast, frugal and primed response (DPT System 1) and has been connected to heuristic, associative, holistic, inferential, and affective dimensions (Hill, 1987; Stanovich & West, 2000; Volz et al., 2008, Stanovich, 2011; Dennin et al., 2022; Svenson et al., 2023). From a sensemaking perspective, research is identifying a range of responses ranging from a heuristic pattern-matching response to a holistic generative response (Glöckner & Witteman, 2010; Chassy & Gobet, 2011; Dreyfus, 2014; Pennycook, 2015). These responses also feature distinct neural processes, with recognition-based responses featuring more engagement of memory areas like the ventro-lateral Pre-Frontal Cortex (vlPFC), Orbito-Frontal Cortex (OFC) and a strong connection with the limbic system (discussed in nuance 3) (Cohen et al., 2005; Poudel et al., 2020; Salehinejad et al., 2021; Colautti et al., 2022; Williams et al., 2025), while insight-related experiences feature different cortical activation around the ventro-medial PFC (vmPFC), parietal lobe and OFC (Volz et al., 2008; Aldous, 2007; Kounios & Beeman, 2009; Andersson et al., 2020). Furthermore, there is evidence for hemispherical differences where the brain's right hemisphere is activated more for remote associations and creativity, while the left hemisphere is more often activated for tighter associations and recognition (Jung-Beeman, 2005; Kounios & Beeman, 2009; McCrea, 2010; Hruska et al, 2015). Other neurocognitive research into the brain's Default Mode Network (DMN) has brought that system in connection with creative ability, coherence-evaluation and generative responses (Beaty et al., 2014; Kühn et al., 2014; Zedelius & Schooler, 2015; Luchini et al., 2025). In summary, the alternative to conscious, rational sensemaking is not merely a heuristic fast lane, but a more complex interplay between

sensorineural inputs, declarative memory (i.e., stored information), procedural memory (brain synaptic "habits") [20] and affect (i.e., emotional tags). Nuance 4 will explore the activation of both heuristic and holistic sensemaking responses.

5.3.1.2 Nuance 2: Interactivity of sensemaking modalities

All three SRK, RPD and DPT constructs describe distinct sensemaking modalities or systems. In extension of the heuristic and holistic responses described in Nuance 1, the brain also features a generally more familiar conscious, rational, analytical response (Rosenbloom et al., 2012; Hruska et al., 2016; Poudel et al., 2020; Hannah et al., 2022). This activates different brain areas including the dorso-lateral PFC (dlPFC) for working memory and executive function, dorsal Anterior Cingular Cortex (dACC) for emotional regulation and insula for salience attention switching (Menon & Uddin, 2010; Collins & Koechlin, 2012; Kühn et al., 2014; Poudel et al., 2020; Salehinejad et al., 2021; Colautti et al., 2022; Williams et al., 2025). The existence of multiple systems has led to (at least) three explanations in how they interact: default-interventionist, parallel-competitive and fuzzy-trace.

The default-interventionist explanation (Evans & Stanovich, 2013) describes a default heuristic (fast and frugal) response to a situation, which can subsequently be intervened upon by an analytical, conscious response if the heuristic response doesn't manifest effectively. This also raises the question whether a meta-system is required to monitor both responses and trigger or inhibit intervention (Nadurak, 2023; Shaw & Nave, 2026). The parallel-competitive explanation (Martin & Sloman, 2013) explains that both intuitive and rational processes co-exist and that the brain chooses the most appropriate response from either approach. A third explanation is the fuzzy-trace theory (Barrouillet, 2011; Reyna, 2012; Blalock & Reyna, 2016) which describes making sense of a situation either through its verbatim "trace" (e.g., text, numbers, graphics) or by a situations bottom-line meaning known as a "gist" (e.g., gain/loss, confidence, sense of uncertainty). Studies show that gist appreciation is often preferred when making decisions, and in some cases may also be the preferred decision-making perspective when facing ambiguity or stochastic extremes (discussed in Nuance 4). Neurological research confirms multi-processing, but rather in a collaborative way rather than a competitive way, often described as a quasi-rational interaction between intuition and insight (Hodkinson et al., 2008; Zander et al., 2016; Calabretta et al., 2017; O'Doherty et al., 2017; de Neys & Pennycook, 2019).

[20] This refers to neurological "procedural memory" i.e. neural structural modification (see Chassy & Gobet, 2011; Dreyfus, 2014; Conway-Smith & West, 2022). Not to be confused with aviation "procedural knowledge".

These studies describe a process where intuitive sensemaking feeds rational insight with novel hypotheses, ideas and constructs (not all plausible, of course). In turn, rational sensemaking provides gaps, queries and problems for heuristic and holistic appraisal. Together, these processes become effective in coherence building (Öllinger & Von Müller, 2017; Akinci & Sadler-Smith, 2020; Van den Berg et al., 2020; Woiceshyn, 2022), which is also supported by brain hemisphere co-activation studies (McCrea, 2010; Hruska et al., 2016; Zamani, 2019; Anderson et al., 2020). In summary, effective and context-responsive sensemaking should appreciate the interaction and value of both conscious, deliberate processes as well as unconscious, tacit ones, without the illusion that one can consciously select between them.

5.3.1.3 Nuance 3: The role of emotions and affect

It is now recognized that both conscious and unconscious mental processes can be influenced by emotions originating from our limbic system (Hogarth, 2002; Dane & Pratt, 2007; Glöckner & Witteman, 2010; Chassy & Gobet, 2011; Salehinejad et al., 2021; Dennin et al., 2022), known as an *affective response* to stimuli. In the domain of executive functioning (EF) i.e. conscious, top-down, deliberate thinking, strong affective decision making is termed "hot EF", which features more limbic engagement through the ventral PFCS and ACC, Posterior Cingular Cortex (PCC), amygdala and insula. In contrast, "cold EF" features less limbic activation and relies more on dorsal and lateral brain regions (Poudel et al., 2010; Salehinejad et al., 2021). Research into such EF shows a strong coupling of hot EF and cold EF with the nature of the situation featuring either risk or ambiguity (Krain et al., 2006; Poudel et al., 2020; Salehinejad, 2021; Colautti et al., 2022). The contexts of reward, value and punishment as well as decisions implying risk (e.g. outcome uncertainty) feature a strong activation of hot EF, such as monetary gains/losses or strong social implications. In contrast, cold EF presents when the situation features no emotional valence, or when experiencing ambiguous situations (e.g. when options are equivalent, discussed in Nuance 4). Affect has also been connected with intuitive appraisal of one's own sensemaking through a metacognitive *feeling of rightness* (Thompson, 2009; McCrea, 2010; Thompson & Morsanyi, 2012), which in turn may also be connected with somatic markers (Ten Houten, 2016) and fluency of retrieval (Volz et al., 2010). How the brain integrates and weighs both internal (e.g. emotional) and external stimuli is still a very active field of research.

Two large meta-studies (Lerner et al., 2015; George & Dane, 2016), exploring 35 years or research in the function of emotions in our sensemaking and decision-making, highlight that emotions can introduce bias, but can also provide useful "metatags" to our stimuli. Critically, the ability to recognize emotions for the role they (can) play is critical. Some explorations into emotional regulation in the field of aviation have focused on recovering from startle and surprise events and strong amygdala responses to mitigate bias

(Field et al., 2018; Landman et al., 2020; Duchevet et al., 2024; Vlaskamp et al., 2025). Research by Brand et al. (2007) indicates the amygdala and limbic responses may even improve decision making under uncertainty, yet this positive contribution of limbic tacit knowledge is still to be explored in the context of aviation sensemaking.

5.3.1.4 Nuance 4: The nature of the situation

The nature of the situation has a profound impact on the activation and effectiveness of various sensemaking mechanisms, both conscious and unconscious. One particular distinction is that between well-structured problems where rules and principles can be applied, and poorly structured problems where no rules or principles have been established and feature significant ambiguity. Through a cognitive lens, the nature of ambiguity can better be understood as *equivocality of options*: an "excess" of certainty, rather than a lack of it (Zeki, 2004; Carleton et al., 2007; Reed, 2015; Julmi, 2019; Stoycheva, 2025). Consider multiple options (e.g. diversion airfields) that have respective pro's and con's in multiple dimensions. Studies show that in such situations, a conceptual, analytical approach does not provide additional sensemaking contrast, and a holistic intuitive appraisal can lead to a better decision (Sadler-Smith & Shefy, 2004; Dane & Pratt, 2007; Dörfler & Ackermann, 2012; Julmi, 2019).

Nuance 3 already introduced another contextual sensitivity for risk and ambiguity in relation to affect. This is furthermore supported by hemispherical differences in the parietal cortex, where straightforward and well-structured problems featured conceptual, left brain activation of regions connected with numbers and comparisons, where ambiguous and less straightforward situations resulted in holistic, right brain activation of regions connected with spatial awareness, visualization and creativity (Krain et al., 2006; Aldous, 2007; Zedelius & Schooler, 2015; Fabio & Towey, 2018; Williams et al., 2025). A possible human factor key to effective ambiguity management is an individual's *tolerance for ambiguity* (Carleton et al., 2007; Stoycheva, 2010; McLain et al., 2015; Stoycheva, 2025). Studies in medical (Iannello et al., 2017) and air defense personnel (Adams-White et al., 2017) highlighted the value of tolerance for ambiguity in contexts that featured regular ambiguity and may prove to be a valuable domain for selection and training of future aircrew.

5.3.1.5 Nuance 5: Differences between experts and novices

A final important nuance in sensemaking is appreciating the interaction between expertise and sensemaking. Studies in the domain of medical diagnostics and chess players show distinct differences between novices and experts (Hodgkinson et al., 2008; Epstein,

2010; Dreyfus, 2014; Klein, 2015) dealing with ambiguity [21]. A notable difference is novice recruitment of mainly top-down conceptual analytical regions and uncertainty estimations (left vlPFC), where experts rely more on holistic intuitive functions, spatial perception, pattern recognition and memory (right dlPFC, parietal cortex) (Hogarth, 2002; Hruska et al., 2016; Hannah et al., 2022; Williams et al., 2025). This is not unsurprising as domain novices do not feature the implicitly learned experiential basis and tacit relational knowledge (De Houwer, 2019) required for heuristic and holistic sensemaking and may therefore best approach problems in a deliberate fashion. There is even light evidence that effective ambiguity management is positively correlated with older age (Zamarian et al., 2008). As such, a differentiation between novice and expert sensemaking may implicate differences in their respective sensemaking contributions, and sensemaking training.

5.3.1.6 Overview of sensemaking archetypes

The five nuances showcase a much more dynamic and multi-factorial playing field that drives which cognitive processes are (or should be) engaged. While the research into these neurocognitive functions is still very much ongoing, a re-appraisal of these functions in the context of complex and ambiguous flight operations provides a more actionable construct for the manifestation of Airmanship 2.0. Table 5.2 proposes three sensemaking neurocognitive archetypes and key factors that distinguish them.

[21] While the game of chess features characteristics of a well-structured problem with clear rules, it is considered an ill-structured problem due to the diversity of both player's options and behaviors, with the number of game possibilities (Shannon number) of around 10^{120} (Newell & Simon, 1972).

Table 5.2. Overview of sensemaking neurological archetypes

Sensemaking neurocognitive archetype	Conceptual sensemaking	Heuristic sensemaking	Holistic sensemaking
Makes sense by	Cause-effect modelling (reconstructing) a complex situation or system	Identifying similarities to other (previous) experiences and situations (schemas)	Immediate appraisal of the situation gist through internal symbology
Mechanism	Conscious, rational ("cold" executive function)	Mostly unconscious, affective ("hot" executive function)	Unconscious, affective, direct stimuli valuation (a-rational function)
Useful for	Hypothesis testing, coordinating a model with other actors	High-risk situations (time pressure, stochastic uncertainty)	Highly ambiguous situations (equivocality of options)
Limitations	Time consuming; limited in resolving ambiguity	Oversimplification of schemas and affective bias risk	Some affective bias risk; expertise required (tacit knowledge)
Relative brain activation	dlPFC; dACC; insula	vlPFC; mPFC; OFC; vACC; PCC; amygdala; insula	dlPFC; vmPFC; ACC; OFC; amygdala; insula; basal ganglia
Developing this requires [22]	Stimuli sensitivity training (allo-centricity)	Emotional regulation training; explicit learning (training scenarios)	Emotional awareness training; implicit learning (tacit knowledge and symbology)

It must be re-emphasized that these are not distinct, selectable modalities of sensemaking, but rather serve to illustrate three zones across a *spectrum of sensemaking*, also highlighted by the close proximity and overlap of cortical region activation. However, this neurocognitive disambiguation of sensemaking mechanisms helps to identify their relative utility in navigating opaque situations. Three key points become clear in this regard:

1. *Conceptual* sensemaking has limitations the face of ambiguity;
2. The possible value of *intuitive holistic* sensemaking in responding to ambiguity; and
3. The influence of *affect* on our intuitive responses.

[22] Training of sensemaking archetypes is addressed in Section 5.4.

5.3.2 Sensemaking neurocognition in the context of cockpit operations

The Airmanship 2.0 concept was incepted to enable effective flight crew responses to the full range of situations, both transparent and opaque, ambiguous and complex. Where procedural responses are effective for transparent situations (which are neither ambiguous nor complex), effectively navigating opaque situations may require a deeper appreciation of the above repertoire of sensemaking mechanisms relative to the nature of the opacity experienced. The relative virtues of conceptual, heuristic and intuitive sensemaking varies across two *dimensions of opacity*:

1. **Ambiguity:** Is the situation/problem well-defined or ill-defined; and
2. **Complexity:** The number of variables and dimensions in the situation/problem.

Figure 5.2 below depicts which responses may be most appropriate based on the nature of the situation. This being a simplification, it should be regarded as a starting point to make sense of sensemaking. The responses will be briefly discussed with an example (from left to right):

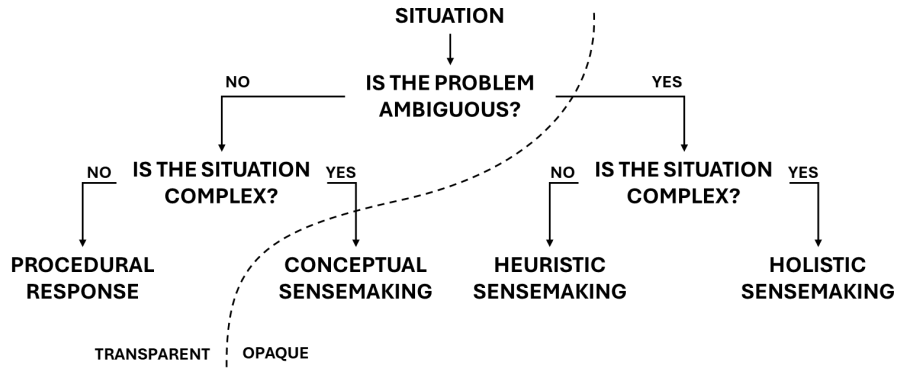


Figure 5.2. Possible classification of neurocognitive sensemaking according to dimensions of opacity

Situations that are not ambiguous and not complex are *transparent*, and as such do not feature any opacity that requires sensemaking. In such situations, the existing deterministic, procedural response is effective without requiring further sensemaking. For example, a single aircraft electrical generator failure where a reset, circumvention or auxiliary replacement procedure can be performed according to the aircraft and operator's standard (non-normal) operating procedures. A situation which is not ambiguous but *is* complex does qualify as opaque due to the existence of multiple concurrent developments (e.g., failures, trends) or contingencies that are not merged into a pre-determined response. For example,

two concurrent generator failures where two or more unprioritized procedures may apply. In this case a *conceptual* sensemaking approach is applicable as the problem is well-defined and allows for rational cognitive modelling. While more rare than transparent situations, this response has already received significant guidance and training in past decades through mnemonic responses such as FORDEC, DESIDE and T-DODAR (Li et al., 2014; Banks et al., 2020).

Further to right in Figure 5.2, situations that *are* ambiguous provide a challenge where multiple explanations or alternatives are cognitively experienced as equivocal. This is not a problem of lack of information, but rather a problem of insufficient contrast (Zeki, 2004; Julmi, 2019). In ambiguous situations which are *not* complex, a heuristic response may already provide a sufficient appraisal of the situation. For example, a generator that drops 30% of output every 18 or so minutes, which could be (heuristically) rationalized as either a failed generator, a surging engine or another intermittent failure and their associated responses. For situations which are both ambiguous *and* complex, holistic appraisal is most appropriate. For example, an intermittently failing generator and an AC-bus failure may give a (gut) sense that the greater electrical system is not to be trusted (an example of a situation's bottom-line *gist*).

The appreciation of different modalities of opacity is important for future flight crew to identify the most effective sensemaking and decision-making approach. A fundamental *openness to opacity* as described in the Airmanship 2.0 concept may be established through the simple appreciation of the modalities presented in this chapter, without a strict "proceduralizing" of the tree presented in Figure 5.1. These sensemaking modalities can only truly be supportive to flight crews when they are recognized for their value in resolving ambiguous and/or complex situations.

All sensemaking modalities have interactions with the limbic system, which can be both a positive and negative force. Conceptual sensemaking (which is mainly a Pre-Frontal Cortex (PFC) activity) can suffer from a strong stress response where PFC activity is reduced. Recognition of this effect and recovery of PFC activity is essential for this sensemaking modality to function and has been a central objective for many startle and surprise recovery training programs. In contrast, heuristic sensemaking taps into "emotional tags" as part of the heuristic knowledge base. As such, regulation of the emotional state becomes important to ensure that a pilot benefits the most from his/her total cognitive capacity (including emotional memory), without becoming too biased by his/her current emotional state. Lastly, the intuitive, holistic sensemaking process also engages emotional memory, but does so in a subconscious, a-rational manner. Similar to heuristic decision making, current emotions may also impact what subconscious patterns and memories drive an underlying "gut feel",

however this can be appreciated less as it does not preside in our conscious brain. For this reason, regulation (and awareness) of both acute and chronic emotional arousal may be essential to maintain the accuracy of all our sensemaking abilities and may be required as an integral trait of the future pilot as proposed in the Airmanship 2.0. concept.

The third important translation of these neurocognitive sensemaking modes to the flight deck underscores their very output: *a novel understanding of reality*. In transparent situations, established mental models guide both awareness and decision making. In opaque situations, these models do not map to the situation, and a revised understanding (e.g., more detailed, rescoped, adaptive) is the result of either analysis, heuristics or intuition. In flight operations, the explicit appreciation that *existing mental models may be limited* is essential to be able to appreciate the output of any sensemaking activity. These model adaptations do not have to be significant and may even be situational (e.g., only this route, aircraft, crew), but they must be admissible inputs to build situation awareness and base decisions on. Both the adaptivity of mental models and openness to opacity underscore the very real possibility of a non-deterministic situation (albeit rare) and would require a pilot's personal tolerance for ambiguity (Stoycheva, 2025). Such openness, tolerance, adaptivity and emotional regulation invites a reappraisal of flight crew training for these sensemaking modalities to function effectively, which is the subject of Section 5.4.

The consolidation of these sensemaking abilities through Airmanship 2.0 does not represent a radical departure from prescriptive, procedural operations but rather proposes an enhanced way of using existing procedural elements. A basic mantra would indeed be: "*the checklist is a tool, not a plan*". Most pilots will acknowledge that this mantra has been part of their training at some point, although pilots will vary widely in their dedication to living by it, depending on their background, experience, attitude, company standard operating procedures and cognitive abilities. It is an illusion that subsequent generations of pilots will be procedure-free and completely unconstrained in their strategies to manage complex and ambiguous aircraft systems and operations. Procedures offer key advantages (e.g., standardization, shared awareness, protection against slips and lapses) and will continue to play an invaluable role in reducing the mental workload associated by partitioning complex interactions of humans, systems and operations into manageable working instructions. However, they may greatly be improved upon if they can be framed and recognized for their sensemaking value, rather than only their executive contribution. The implication of this way of working is that the pilot no longer faces the existential "*do I comply or do I deviate*" role bifurcation. Instead, he/she may comply with procedures within a (meta-)process of learning about and resolving situations. An important implication of this is that "deviations" no longer exist as an error mode: procedures are no longer the operational gold standard, they

are (merely) tools for the pilot to use. This shift in standards will briefly be discussed in Section 5.6.

Taken at face value, the Airmanship 2.0 nested way of working would seem to suggest potential increase in mental workload. Rather, it underscores the value of open-awareness and intuitive salience sensitivity (coupled with a fail-safe *mindset*) as low workload "canaries in the coalmine" (Menon & Uddin, 2010). Mental workload will only increase if indeed the situation is requiring an activation of conceptual sensemaking which requires considerably more working memory (Andersson et al., 2020; Salehinejad et al., 2021).

5.4 Training sensemaking

This deeper apprehension of the neurocognition of sensemaking invites a re-appraisal of how pilots are trained to develop competence across all three sensemaking archetypes. This section will differentiate between novice and expert pilots in line with Nuance 5 and as well as argue the necessity of novice-expert teaming for diversity of sensemaking mechanisms and – critically - ensuring continuity in workforce expertise.

5.4.1 Training novice pilots

Novice pilots may be the subject of two evolutions in training: the first is expanding the *general professional foundation* of a pilot with a tolerance of ambiguity and emotional self-regulation which is carried forward through their career. The second is novice-specific neuro-cognitive training to enhance conceptual sensemaking by increasing *external* stimuli awareness and reducing *internal* limbic and self-referential inputs (which can be considered "noisy signals" when such internal inputs aren't sourced from a broad foundation of experience, expertise and knowledge).

As novices lack the tacit knowledge and experience to solicit expert intuition, they may best engage a conceptual, top-down rational approach to sensemaking. At the neurocognitive level, four distinct training objectives can be identified for novice pilots:

1. A basic tolerance for ambiguity (openness to opacity) to appreciate the value of sensemaking;
2. Down-regulation of emotions to reduce self-referential "noise" and maintain Pre-Frontal Cortex (PFC) working memory;
3. Increased attentional control for sensitivity to external stimuli, improving the quality of inputs to the sensemaking process;
4. Fluency in conceptual sensemaking techniques contextualized for flight operations;

An important second-order effect of the above requirements may be a greater resolution and quality of the developed tacit knowledge and internal schemas (Talbot, 2004, Klein, 2015) that result from a richer experience of stimuli – i.e., a sharper intuition. This is essential to ensure availability of expert intuition in the future.

A promising avenue for novice neurocognitive training is mindfulness training, which may be understood as a the collection of deliberate practices in mental control and though awareness. Mindfulness proficiency potentially addresses all four training objectives. Two main mindfulness practices are explored in this chapter: focused attention mediation and open awareness meditation. Developing mindfulness through *focused attention meditation* practice has been shown to reduce self-referential thinking (training objective 2) by reducing Default Mode Network (DMN) activity (Farb et al., 2007; Brewer et al., 2011; Marchand, 2014; Garrison et al., 2015; Tang et al., 2015; Li et al., 2020) and improve allo-centric attentional control to external stimuli (training objective 3) (Pagnoni et al., 2008; Hölzel et al., 2011; Garrison et al., 2015; Oken, 2016; Bremer et al., 2022). The same attentional control mechanism also improves emotional awareness (training objective 2) and in turn emotional self-regulation (Hölzel et al., 2011; Marchand, 2014; Tang et al., 2015; Mansouri et al., 2025). Mindfulness has also been connected with a faster post-stress deactivation of the amygdala (such as after a startling or surprising event) (Brewer et al., 2011; Hölzel et al., 2011; Meland et al. 2015, Alaydi et al., 2025). From a sensemaking perspective, this improves accessibility to cold executive functioning (EF) and attenuates hot EF and intuitive functions (training objective 4). There is even evidence that mindfulness improves creative problem solving through *analysis* (Zedelius & Schooler, 2015; Berkovich-Ohana et al., 2017; Byrne & Thatchenkery, 2019). Furthermore, mindfulness training may also develop a novice's openness to opacity (training objective 1) by reducing a need for control (egocentric response). Evidence of this is found in the context of Zen training where training in ambiguity acceptance is a central tenant through concepts such as *wu-wei* ("not-doing") and koan training (Brezina, 2011; Low & Purser, 2012; Austin, 2013; Li, 2014).

Beyond the cognitive foundation provided by mindfulness, effective conceptual sensemaking requires integration into normal operations to avoid its atrophy (Mohrmann, 2026). This could possibly be achieved by adapting existing aviation-relevant problem-solving mnemonics such as FORDEC, DESIDE and T-DODAR as vehicles to operationalize conceptual sensemaking on the flight deck (training objective 4). As the high reliability and standardization of modern flight operations provides infrequent opportunities to develop sensemaking in operations, training programs for novices must feature a large amount of context-rich and (progressively) unguided training to practice and gain confidence in their sensemaking abilities, be tolerant of ambiguity and grow their repertoire of solutions and internal schemas. An example of such exploratory, unguided training was recently

explored by Van Leeuwen et al. (2024). This type of simulated training can be supported by on-the-job development of expertise through *deliberate performance* (Fadde & Klein, 2010). Such training captures job experience explicitly as expertise-building opportunities to build better mental models, rather than only relying on implicit learning. This concept strongly mirrors the concept of *work-practice* from the Zen school of meditation. The Zen school emphasizes developing mindfulness through direct, experiential meditative practice rather than active mental training. In the domain of Competency Based Training & Assessment (CBTA), this is often referred to as developing *self-reflecting practitioners*, where (flight) operations are regarded as stages to actively learn and develop, rather than only relying on formal training contexts.

5.4.2 Training expert pilots

Expert sensemaking training proposes building on the above foundation of novices by assuming a existing tolerance for opacity, proficiency in emotional regulation, conceptual sensemaking and (growing) tacit knowledge and experience base. Training experts should focus on developing proficiency in heuristic and holistic sensemaking to manage particularly ambiguous situations, addressing the following training objectives:

1. Increase the expert's aperture to intuitive responses and implicit knowledge;
2. Improve creativity and remote association of concepts;
3. Fluency in heuristic and holistic sensemaking techniques contextualized for flight operations;

Training familiarity with intuitive responses requires tuning into implicit, internal knowledge including affective, somatic and a-rational experiences (i.e., insight, or "aha-erlebnis"). This is a departure from the novice pathway which was focused at quieting such bottom-up internal signals in favor of external stimuli and top-down rationalization. Neurologically, this implies a more active Default Mode Network (DMN) for improved creativity in response to ambiguity, by means of more active associative thinking (Volz et al., 2014; Berkowich-Ohana et al., 2017; Fabio, 2018; Luchini, 2025). Remmers et al. (2015) also hypothesize that a slight reduction of cognitive control permits a modicum of judgmental behavior required to be able to attribute value to bottom-up intuitive insights. Associative thinking does require a sufficiently mature tacit knowledge base to associate with and as such should be reserved for expert training. Furthermore, studies linking long-term mindfulness practice with creativity also show that such practice induces a shift from left-brain rational to right-brain holistic thinking and features greater cognitive flexibility (Manna, 2010; Fabio, 2018; Luchini, 2025). This is considered generative (associative) creativity, rather than the analytical creativity that novices may experience.

The transition from novice to expert sensemaking introduces a shift in the type of mindfulness meditation practice required. Where novice training emphasizes developing cognitive control through *focused attention*, expert training should emphasize intuiting and sensing through meditation practices focused on *open awareness* (Berkowich-Ohana et al., 2017; Fabio & Towey, 2018). In Zen training, this is also described as direct experience (Pagnoni et al., 2008; Austin, 2013), which has for instance historically been trained through practices with Zen koans which cannot be solved rationally (Low & Purser, 2012; Li, 2014). Neuroimaging of both novice and expert Zen monks shows distinct differences where most notably open awareness meditation shows activation of both left (analytical) and right (holistic) brain regions (Manna et al., 2010), echoing similar findings observing brain activity during expert medical diagnosis of ambiguous situations (Van den Berg et al., 2020). The practice of open awareness implies, in contrast to focused attention, a relaxation of the mind. Interestingly, this aligns with experiential research showing that insight (i.e., "aha erlebnis") arises when one is in a state of relaxation such as under the shower, where right brain activity (holistic, associative) was observed prior to experiencing insight (Ovington et al., 2018).

Training such expert sensemaking requires sufficient exposure to situations that allow a candidate to experience ambiguity (in the context of aviation), practice open awareness and navigate the cognitive experiences of heuristic and holistic sensemaking. In contrast to novice sensemaking training focused on building tacit knowledge, expert training situations may feature a greater level of ambiguity and complexity that are far less proximal to familiar or even realistic situations. Where traditional pilot training emphasizing manual skills, procedural compliance and declarative knowledge is thereby limited to realistic situations (i.e. situations that can be expected to occur), sensemaking training may venture deeper into increasingly unlikely scenarios – a proverbial aviation *koan*. Such situations may also echo accident cases which featured significant ambiguity such as British Airways flight 009 that flew through volcanic ash, Quantas 32 which experienced 53 malfunctions [23] or Flight 1549 which featured a double engine failure shortly after departure. This may also develop an expert's intuiting through *creative action* where intuitive sensemaking is stimulated through unconscious somatic and sensorineural engagement with the aircraft, system or situation (Caruthers, 2007; Berkovich-Ohana et al., 2017; Stoycheva, 2025). In other words, taking action activates deeper sensemaking responses. A great example of this is showcased in Quantas Flight 72, in which a pitch control system reversal problem was only identified through the captain's intuitively *letting go* of the flight controls (ATSB, 2011). This is a similar experience as learning to solve a Rubik's cube through manipulation rather than solely through conceptualization, trimming sails when sailing with gusty winds or driving a car in

[23] 53 malfunctions indicated by the A380's Electronic Centralized Aircraft Monitoring (ECAM) system.

busy traffic – our interactions with system and the environment *feed* our understanding as much as they are driven by it. As such, it is not only useful but may even be imperative that flight crews regularly interact with the aircraft and systems, in contrast to a role centered heavily on monitoring and automation management.

5.4.3 Novice-expert sensemaking teaming

The neurological differentiation between novice and expert pilot contributions to sensemaking offers a new perspective to the flight deck JCS. While an expertise gradient exists between captains and first- and second officers, the current interchangeability between pilot flying (PF) and pilot not flying (PNF) roles may need to factor in what sensemaking strategy the crew can best employ. For example, if a particularly ambiguous situation requires expert intuition, the more experienced crew member may need to engage in a diagnostic PNF role (or PF role, if the situation features flight controls problems in particular). The most experienced crew member would be the captain in most cases but may also be a first officer that has more experience on the aircraft type in question (Weick & Sutcliffe, 2011). In contrast, less ambiguous situations provide opportunities for less experienced crew members to response either procedurally or conceptually to develop their tacit knowledge through deliberate performance. Such novice-expert teaming has been suggested to be critical in developing expert intuition (Dane & Pratt, 2007).

Another crewing dynamic that may be relevant is operating a dual process approach with one crew member sensemaking conceptually, and the other sensemaking intuitively. This mirrors the brain's interaction between hemispheres, where conceptual analysis provides stronger diagnostic hypotheses and intuitive responses provide novel explanations for ambiguous and complex situations. Such a collaboration could be the subject of new neurologically informed problem-solving mnemonics that engage specific sensemaking strategies. This may be a way to systematically maintain sensemaking fluency in light of the comparatively low occurrence of ambiguous situations as proposed by Mohrmann (2026). Such an approach mirrors the harmonization of organizational mindfulness (Slagmolen et al., 2009) into currently hierarchal sociotechnical systems as proposed by Lintern & Kugler (2017).

Such advanced neurocognitive crew collaboration for systemic resilience to opacity introduces significant implications for the recent flight deck evolutions towards Reduced Crew Operations (RCO) (Malik & Gollnick, 2016; Sprengart et al, 2018; Schmidt & Stanton, 2020) and Human-Artificial Intelligence (AI) Teaming (HAT) (EASA, 2023; Lorrig & Daw, 2024; Kirwan 2025; Hörhager, 2025) which, respectively, remove and add to the cockpit cognitive space. The concept of RCO is difficult to reconcile with the above neuro-collaborative paradigm. This is because it either inhibits progressive expertise development

(required for resilient operations) or significantly raises pilot expertise requirements which may result in a workforce continuity challenge (Harris, 2023). The possible expansion of supportive automation in the RCO pathway also risks a stronger re-activation of Klein's self-reinforcing complexity loop (Klein, 2011), which is the very *raison d'etre* for a more resilient Airmanship 2.0 concept. The emergence of AI in aviation also possibly introduces ripples in the cockpit JCS and may either exacerbate or mitigate systemic brittleness of the Flight Deck Socio-Technical System (FDSTS). Novel human-AI teaming protocols which actively facilitate the human neurocognitive sensemaking mechanisms described in this chapter may indeed help reposition the future pilot toward Airmanship 2.0. Both RCO and HAT developments may prove to be critical topics for future research into sensemaking.

5.4.4 Possible evolutions of the training pipeline

To support the previously established neurocognitive differentiation between novices and experts, pilot training and development can/must exhibit a phased trajectory to provide the right sequence and separation in training these problem-solving skills over the course of a pilot's career. Table 5.3 shows the current training pathway for a commercial pilot and indicates how these programs could be modified to develop the neurocognitive concepts underpinning Airmanship 2.0. An expanded reconsideration of the training pipeline is provided in Appendix H.

Table 5.3. Overview of possible adaptations to the training regimen of a commercial pilot for sensemaking

(Current) Training Program	Possible adaptations
Pilot selection	▪ Evaluation of key sensemaking pre-requisites including self-reflection, basic emotional regulation skills and tolerance for ambiguity;
Ab-initio & Multi-Crew Coordination (MCC) training	▪ Mindfulness training (focused attention meditation); ▪ Emotional regulation training (e.g. startle and surprise recovery training); ▪ Scenario-based conceptual sensemaking training; ▪ Crew-coordination training in collaborative sensemaking strategies;
Type-Rating & Zero Flight Time Training (ZFTT)	▪ Practice type-specific conceptual problem solving;
Recurrent Training	▪ Recurrent training in sensemaking and emotional regulation (e.g. recurrent startle & surprise training); ▪ Enhanced CRM training in collaborative sensemaking; ▪ Adapted competency model: ▪ Extending Leadership & Teamwork (LTW) and Situation Awareness (SAW) competencies with behaviors about metacognitive awareness; ▪ Extending Problem Solving & Decision Making (PSD) competency with behaviors related to cognitive flexibility and hypothesis testing; ▪ Practice in flight debriefings and *deliberate performance;*
Command Training	▪ Selection criteria should assess a pilot's abilities in conceptual sensemaking and potential for heuristic/holistic sensemaking; ▪ Training in heuristic and holistic (intuitive) sensemaking; ▪ Training in experiential learning (creative action);
Instructor training	▪ Ability to teach (collaborative) sensemaking activities; ▪ Facilitation techniques specific to startle, mindfulness, and all three sensemaking modalities;
Examiner training	▪ Ability to evaluate (collaborative) sensemaking activities (ideally competency-based);

Several of the above changes are already being supported in the form of (voluntary) CBTA such as Evidence Based Training (EBT) for recurrent training and the Multi-Pilot License (MPL) program for ab-initio development (ICAO, 2013; IATA, 2024b; EASA, 2020). The coming years will also see CBTA developments trickle down the training pipeline from instructor and examiner qualification, to type-rating training through to (revised) ab-initio training, providing more opportunities for sensemaking training.

These developments provide a good foundation to introduce these new neurocognitive abilities for two reasons. First, competency-based training programs make use of more operationally relevant and realistic training with simulated Line-Oriented Flight Training (LOFT), thereby shifting the focus from *checking* to *training*. Second, they already feature significantly expanded performance criteria to include non-technical crew collaboration skills, which can be further expanded to fully encompass all the (meta-)cognitive abilities described in this chapter.

The following section describes a study exploring these new neurocognitive frontiers in aviation training. As an example of bridging these neurocognitive abilities with the pilot competency framework, the case study in Section 5.5 explores the potential of mindfulness training to improve flight crew response to opaque, ambiguous situations in flight, as well as the challenges in pushing the envelope in training toward the neurocognitive domain.

5.5 Case study: Pilot mindfulness training

This section will outline the case study approach, results and research conclusions. Section 5.6 will reflect on the related challenges (and opportunities of neurocognitive training in aviation. The author was engaged to guide a large, western air cargo operator's safety department in their exploration of ways to develop a High Reliability Organization (HRO), and in particular to identify opportunities to train their personnel in resilience behavior. A key element of HRO development was introducing mindfulness into the organization to improve organizational resilience. The use of mindfulness to support HRO's has been explored earlier (Slagmolen et al., 2009; Weick & Sutcliffe, 2015). The study was performed in 2018 with the knowledge of the time. This specific study explored whether mindfulness training could equip pilots to improve their flight safety performance on the following top safety priorities (at the time, unranked):

- Altitude deviation
- ATC complications/ communication event
- Incursion/excursion event
- Speed deviation
- Navigation/position control
- Non-compliance with regulations, policies, procedures
- At-risk landings
- Unstable approaches
- Automation modes on go-around and landing

The mindfulness training intervention was aimed at developing pilots in the following neurocognitive skills:

- Emotional regulation,
- Non-judgement,
- Observing (focused attention),
- Open awareness (monitoring more sensory info/emotions/thought), and
- Cognitive flexibility (task/situation/concept switching)

To link these neurocognitive abilities to flight safety priorities, the study used a behavioral framework to bridge neurocognitive ability to flight safety. This framework employed the eight [24] ICAO core competencies, supplemented with "startle & surprise management" and "fatigue management" as two custom (non ICAO) competencies, with their respective behavioral markers established by the research team. The choice for this compound behavioral framework was based on the most likely connections between flight operational safety and mindfulness training impact, as well as aligning to relevant industry training standards and human factors research. Each of these competencies was evaluated in its ability to bridge mindfulness with flight operations. This was done by ranking their (expected) response to mindfulness training, and in turn their (expected) impact on the safety priorities.

The first comparison linking competencies and their behavioral markers with *mindfulness training* used a Delphi-approach (Linstone & Turoff, 1975) involving three CBTA and mindfulness experts, to ascertain which competencies and respective behavioral markers would be most positively affected by mindfulness training. The Delphi method overlays independent expert assessments, contrasts responses and drives for consensus where differences occur. The ranking is summarized in the "mindfulness ranking" column in Table 5.4. As this score was based on evaluating each competency's behavioral markers on potential mindfulness impact, final scores were normalized against the different number of behavioral markers in each competency. The larger the percentage score, the greater its potential to be positively influenced by mindfulness training. The second comparison linking the competencies and their behavioral markers with *flight safety priorities* was performed via the Delphi method involving three CBTA, safety and operational experts. The results are

[24] This study engaged the eight ICAO core competencies in line with ICAO Doc 9995 (ICAO, 2013) which was the recognized pilot competency framework at the time. In recent years, the industry recognizes a distinct ninth competency ("Application of Knowledge") which was introduced by EASA and IATA and is now standardized with ICAO.

provided in the "Safety Ranking" column in Table 5.4. The "Average ranking" column is the average of the two other columns and indicates which behaviors serves as the most robust bridge between mindfulness and the safety priorities and serves as the rank order for the table.

Table 5.4. Relationship of mindfulness to flight safety priorities via competencies (ranked by average rating)

Competency (*not ICAO Competency)	Mindfulness ranking (Normalized)	Safety ranking (Normalized)	↓Average ranking (Normalized)
Problem Solving and Decision Making	100%	100%	100%
Workload Management	96%	95%	95%
Startle & Surprise Management*	98%	85%	92%
Situation Awareness	94%	81%	88%
Aircraft Flight Path Management (FPM) – Manual	80%	92%	86%
Aircraft FPM – Automatic	76%	91%	83%
Leadership and Teamwork	85%	81%	83%
Communication	86%	69%	78%
Application of Procedures	49%	85%	67%
Fatigue Management*	40%	91%	65%

Based on Table 5.4, the main research hypothesis to connect mindfulness and flight safety focused on those three competencies featuring the greatest connection to both factors (highest "Average Ranking" according to Table 5.4, i.e. the top three competencies), and was stated as follows:

> *"Mindfulness training is expected to have a positive effect on the operator's flight safety through supporting and enhancing the following flight operational behaviors/core competencies: Problem Solving & Decision Making, Workload Management, Startle & Surprise Management."*

5.5.1 Study method

The study explored this hypothesis by comparing flight safety performance between two groups: a control group without mindfulness training and treatment group with mindfulness training. The treatment consisted of two months of mindfulness training prior to a performance evaluation in a simulator scenario. The control group was not offered a placebo training; they represented the current state of the operator's flight crew proficiency.

Each group initially consisted of 15 crews (30 pilots), all from the same multi-crew long haul fleet, with the two groups balanced as much as possible for rank, age and flying experience.

5.5.1.1 Mindfulness training and evaluation scenario

The mindfulness training was developed by the mindfulness training organization and was tailored toward the operator's flight operations, in collaboration with the operator's flight operations experts. This tailoring explicitly linked mindful behavior with challenges and experiences in and around the flight deck context.

The mindfulness training consisted of two parts. The first part was a face-to-face training day with mindfulness trainers at the operator's training center. This served as a kickoff to ensure that mindfulness practices were learned correctly at the onset. The second part was a series of online coaching sessions to check-up with candidates and reinforce the training for a period of two months. This was supported by a biometric feedback device, the Stanford-developed Spire Stone (Miller, 2017). The Spire Stone was a small, wearable device clipped to clothes around the torso, and measured a candidate's diaphragmatic breathing depth and frequency with a pressure sensor, as well as physical activity with an 3D accelerometer, and distinguished a measure for three mental states: calm, tense and focused using algorithms developed by Stanford University's Calming Technology Lab (Miller, 2017). Through a Bluetooth connection with their phone, candidates in the treatment group received alerts and relaxation tips when tense breathing was detected. The Spire device was provided to all candidates in both test groups to collect and compare biometric data between the groups. The alerting functionality was only enabled for the treatment group to support mindfulness training effectiveness for the treatment group. Besides not receiving training, the provision of biofeedback was disabled to pilots in the control group to minimize/avoid bias and self-training in that group. After completion of all simulator performance evaluations, the control group was offered the same mindfulness training, as well as the uninhibited Spire Stone device.

Safety performance was evaluated in a simulated flight scenario designed to challenge the bridging behaviors in Table 5.4, with a strong emphasis on *Problem Solving & Decision Making*, *Workload Management* and *Startle & Surprise Management*. The scenario featured a short night-time repositioning flight of a wide-body aircraft from the operator's base to a nearby airfield about 300 nautical miles away. The following is a brief synopsis of the scenario. Figure 5.3 shows a visual representation of this scenario.

> *The scenario begins as most flights with this operator: the aircraft is parked at the gate and shut down. The crews are provided all the regular preflight briefing material including aircraft state, weather and routing. They have about 30 minutes to review the paperwork, which is standard. Weather at the departure field is 5 miles of visibility, overcast at Flight*

Level (FL) 100 with cloud tops at FL260 with some thunderstorms brewing underway (not unusual for this region). The aircraft features a failure of the Hydraulic System Controller, which has an "inoperative" sticker on the left side of the hydraulic controller. This item does not restrict the aircraft as per the Minimum Equipment List (MEL). The crews receive a last-minute departure runway change as an initial mild disruption (mild surprise). Normal departure, climb and (short) cruise. Weather at the destination airfield has deteriorated significantly, with several runways closed for landing due to a storm cell that is passing, the crews are "forced" to divert to another more suitable airfield a bit further away. The wind at this alternate airfield is 170/15 gusting 22, cloud base at 3,500ft and 7 miles visibility. There are towering cumulus clouds in the area. Upon descending into the diversion airfield, the flight is cleared for the runway featuring 90 degrees crosswind at 15 knots gusting 22 knots, close to the aircraft crosswind limitations. At 2,400ft altitude during the approach, the anti-skid fails for one of the landing gears (another hydraulic system failure), which requires a go-around given the cross-wind conditions (a surprise event). Missed-approach procedures require them to turn left, to the North. If crews elect to continue on the runway heading instead of turning, ATC will request this for traffic avoidance reasons. After this initial turn, crews turn further left for a second approach, and are then struck by lightning (a startle event, during a higher workload situation). Autopilot and auto-thrust systems disconnect as a result. Crews are given time to react and plan as they see fit, and are provided radar vectors if requested. Crews may elect to choose runway 9L as it is longer and more appropriate for the cross-wind conditions. The remaining approach, landing and taxi-in are uneventful to end the scenario with a successful flight (to avoid negative training).

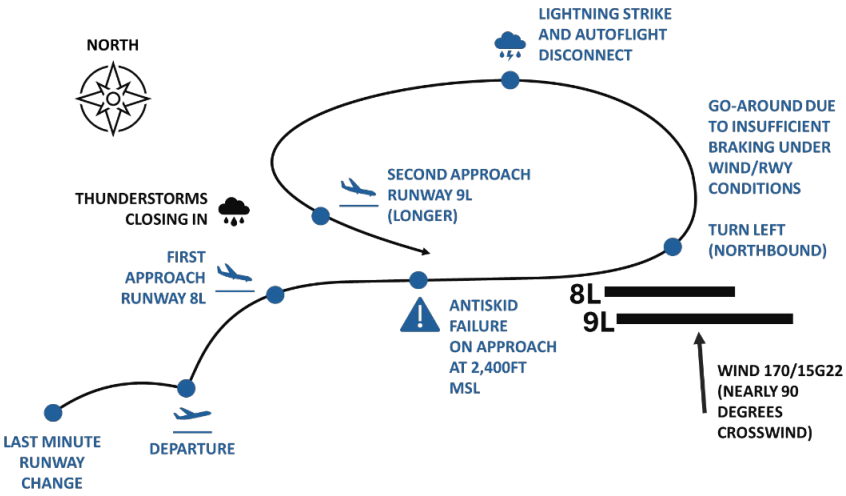


Figure 5.3. Sketch of flight simulation scenario

The scenario duration was approximately 45-60 minutes. To maximize immersion, crews were requested to arrive in uniform, perform all normal briefing items, Standard Operating Procedures (SOP) and check items as they would normally do in an actual flight, and to not treat the simulator as a training opportunity. There was one simulator instructor present to operate the simulator and scenario events, but he/she did not intervene with crews at any point. Despite this scenario featuring many more crew challenges than a normal flight, great care was taken such that the slow escalation, temporal spread and specific timings of events were such that crews were drawn in (immersed) and not bluntly overloaded. Several validation sessions including pilots not involved in the scenario design facilitated the fine tuning of this scenario and event timings.

5.5.1.2 Mindfulness data

The study collected two types of data: mindfulness measures (Subsection 5.5.1.2) and flight safety performance measures (Subsection 5.5.1.3). Mindfulness data were collected at the beginning of the study and during the treatment period of two months, to determine whether the treatment group featured a different level of mindfulness (post-treatment) than the control group. Flight safety performance data were collected to determine if the treatment group also outperformed the control group in the flight operational scenario. All data were collected in a double-blind fashion to avoid evaluator bias.

Data on mindfulness consisted of self-reported scores via two surveys and biometric data from the Spire Stone device, collected for both control and treatment groups. Self-reported scores were based on an 86-question survey administered before the mindfulness training and immediately after the simulator session (Survey 1). A shorter 31-question survey was administered after each actual operational flight the crew member flew during the two-month training period tracked the development and application of mindfulness techniques in flight operations (Survey 2). Both surveys can be found in Appendix I.

Surveys asked candidates to what extent they recognized themselves in specific statements or had particular experiences related to the following five mindfulness measures:

- Focus (Scale 1-10),
- Communication Quality (CQ) (in the flight deck) (Scale 1-10),
- Emotional Intelligence (EI) (Scale 1-5),
- Fatigue (Scale 1-5), and
- Work-Life Balance (WLB) (Scale 1-5)

The biometric mindfulness data provided insights into deeper changes to candidate mental states over the course of the preceding two months. Participants were encouraged to wear the Spire device as much as possible, although it was ultimately voluntary. Data was collected about device usage (minutes per day) as well as the ratio of specific inferred mental/physical states during device usage. Each ratio indicates the percentage of measure time the individual was in a given state The biometric data consists of the following metrics:

- Total minutes worn (group total minutes per day),
- Focus ratio (group average ratio per day, same for other ratios),
- Calm ratio,
- Tense ratio,
- Activity ratio, and
- Sedentary ratio

5.5.1.3 Flight safety performance data

Flight crew performance data were collected from observations of crew behavior during the simulated scenario, supported by facilitated debriefings after the simulator scenario. The Desired Flight Crew Performance (DFCP) method (Field et al., 2016) was used to provide a fair basis of comparing multiple alternative crew resolution strategies in this scenario. The DFCP framework for this study was developed in collaboration with the operator's flight operational experts and human performance experts from the participating research institute. Each performance item was unambiguously described and marked as "observed" or "not observed". Each performance item was phrased as a positive, desired behavior, so marking it as "observed" results in a higher valuation of that crew's performance. By structuring observations and avoiding SME interpretations and individual performance standards, the effect of subjectivity is reduced considerably.

The DFCP framework featured 53 observable items, which were graded "live" during each scenario run, by a type-rated examiner. Only a small group of examiners were engaged, who were also either involved or briefed about the DFCP evaluation format to provide a level of assessor concordance. Post-scenario debriefing was also facilitated by the same examiners and provided further resolution to simulator observations concerning flight crew awareness, decisions, self-awareness, experiences and considerations. These debriefings were recorded and transcribed to supplement a crew's DFCP scores with context and storyline. The observing examiner did not know whether the crew was from the treatment or control group to avoid study bias.

Crew DFCP scores were further divided per flight phase and per competency (across the entire scenario). The division per flight phase allowed the data analysis to investigate whether the experiment groups featured different performance in any specific segments of the scenario (which may feature specific challenges). The division per competency was an alternative approach to determine if there were between group differences in DFCP items related to one competency or another. The distribution of the 53 DFCP items across the study's competency framework was determined with the support of three independent SME's: an active-duty pilot, a CBTA expert and a type-rating examiner. In this cross-reference, the Fatigue Management competency was not attributable to any DFCP item, explainable by the fact that it was also not one of the design focus points for the simulator scenario. Hence, in the DFCP results there is no fatigue competency measure presented.

5.5.2 Study results

Unfortunately, the research team experienced significant (and unexpected) challenges in retaining participants for the full duration and scope of the study. The explorative and progressive nature of this research initiative was difficult to place in both the operator's operational and training contexts. In particular, the organization's messaging about the study rationale and desired outcomes turned out to negatively affect pilot interest and motivation to participate. Many crews disengaged before the final performance evaluation in the simulator, with a few cases completing the simulator flight but whose results were not admissible due to specific union-company agreements. This is a key learning opportunity for future attempts in such neurocognitive and emotionally sensitive training. An overview of responses and participation is provided in Table 5.5. Instead of the intended 30 crews (60 individuals), only nine crews completed the entire experiment including the simulator scenario (five treatment group crews and four control group crews).

Table 5.5. Response rate for mindfulness and performance measures

Measure	Control group	Treatment group
Survey 1 pre-study	23 individuals	32 individuals
Survey 1 post-study	4 individuals	7 individuals
Survey 1 both pre- and post-study	2 individuals	2 individuals
Survey 2 (operational)	59 submissions 12 individuals	161 submissions 29 individuals
Spire biometrics	Varying number of individuals, declining usage	Varying number of individuals, declining usage
DFCP	4 crews (8 individuals)	5 crews (10 individuals)

5.5.2.1 Mindfulness survey results

Initial mindfulness scores from Survey 1 indicated an equivalent level of mindfulness for both control and treatment groups prior to treatment. However, due to the rapid decrease in study participation, evaluating the longitudinal training effect with pre- and post-study Survey 1 responses was not feasible due to the very low response rate to the exit-survey: only four individuals (two in each treatment group) completed both pre- and post-surveys. As such, Survey 1 results are not investigated further in this study.

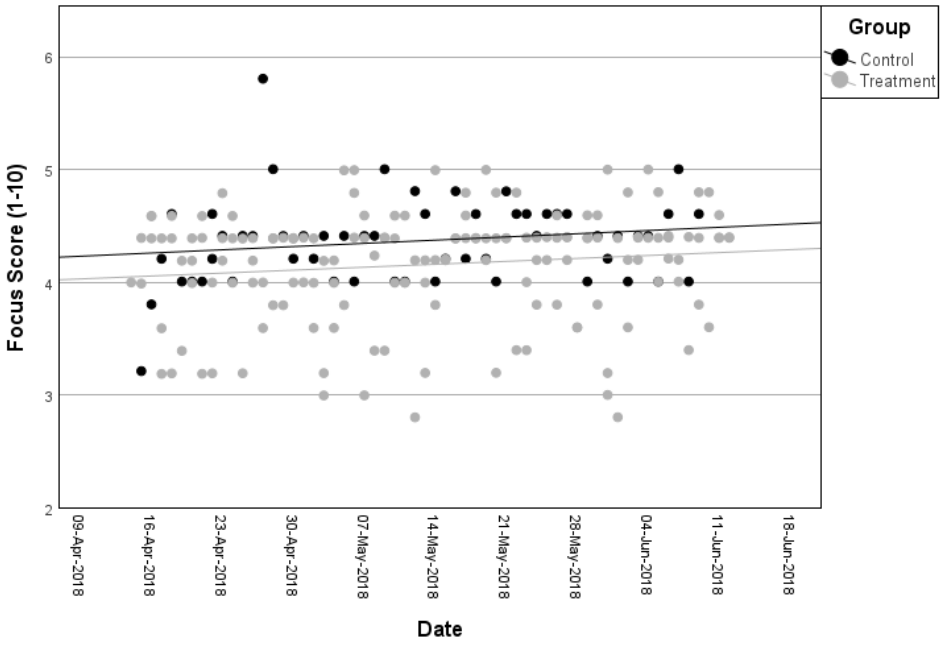
Survey 2 results did have sufficient data to perform between-group evaluations. A very small number of extreme outliers were removed. Shapiro-Wilk testing indicates these metrics should be treated as non-parametric. Non-parametric Mann-Whitney U Test results for between group differences are provided in Table 5.6. A Benjami-Hochberg False Discovery Rate (FDR) *p*-adjustment is applied to each test family and critical *q* is reported as such. Statistically significant results ($q \leq 0.05$) are indicated, as well as near-significant results ($p \leq 0.05$ but $q > 0.05$), and strong trends (p>0.05, effect size >0.5). Furthermore, to investigate temporal between group differences (i.e., changes of scores over time), an Analysis of Covariance (ANCOVA) significance of the group interaction term is also reported. CQ and EI ratings are statistically significantly higher in the treatment group, although there is no difference between the groups over time. Fatigue and work-life balance scores are not statistically different between the groups. Figure 5.4 shows the survey response data plots and Figure 5.5 shows the respective box plots.

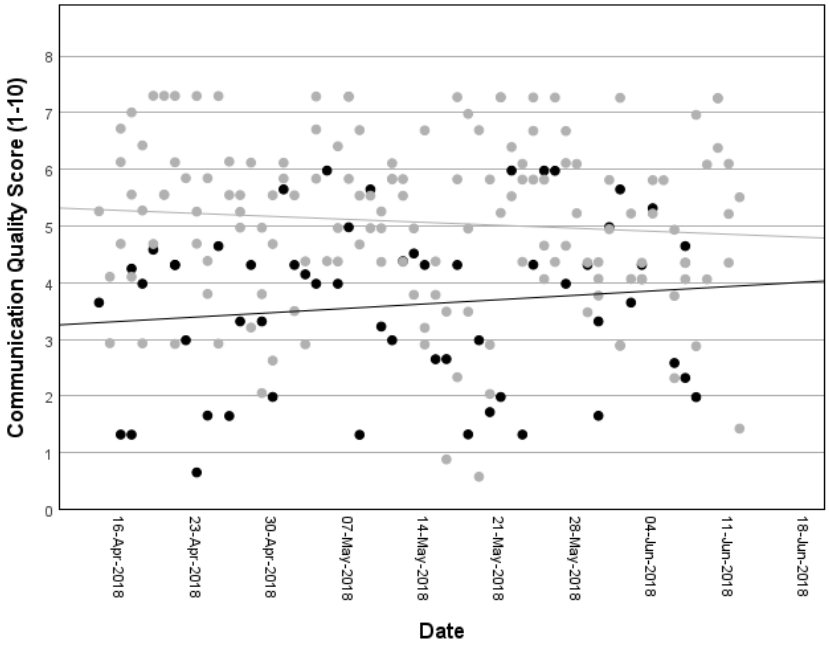
Table 5.6. Survey 2 between group differences (Mann-Whitney U Test)

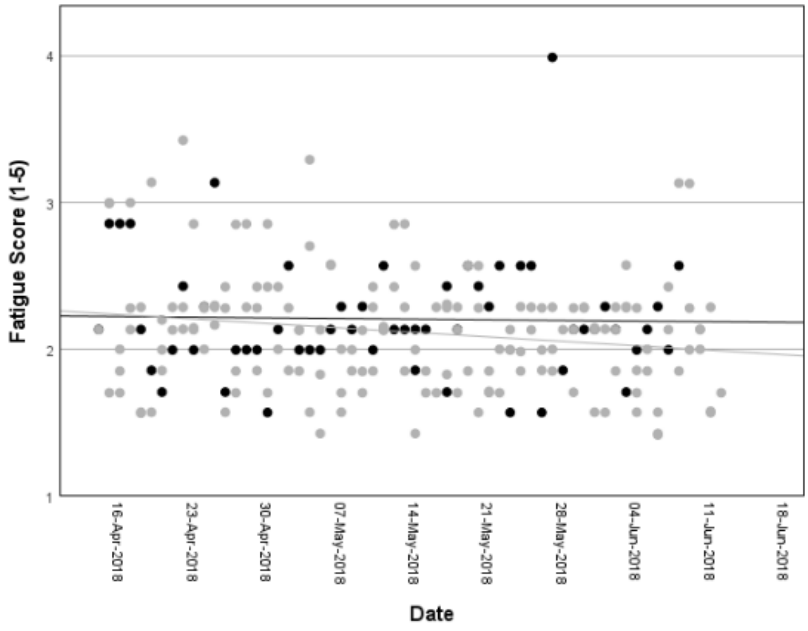
Metric	Control group mean rank	Treatment group mean rank	Exact [25] significance p (two-sided)	ANCOVA significance p
Focus	131.23	96.88	<0.001***	0.142
Communication Quality	64.00	117.66	<0.001***	0.001***
Emotional Intelligence	84.68	113.07	0.003***	0.402
Fatigue	119.35	103.20	0.092	0.866
Work-Life Balance	97.43	113.67	0.096	0.435

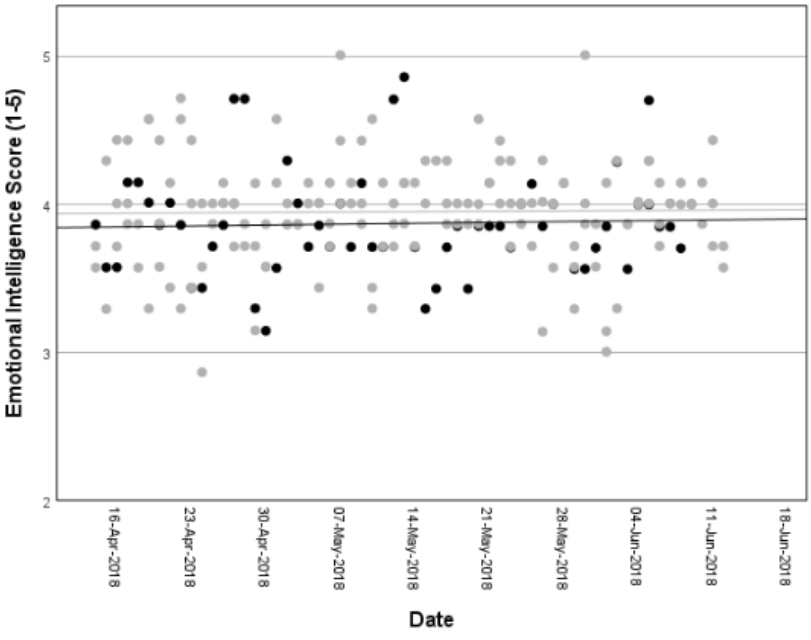
*** $q \leq 0.05$, $p \leq 0.05$, *$q > 0.05$*, *$p > 0.05$*, *effect size $r > 0.5$*

[25] Customary for Mann-Whitney U-Testing with sample sizes smaller than 20.


Group
Control
Treatment
Focus Score (1-10)
6
5
4
3
2
09-Apr-2018
16-Apr-2018
23-Apr-2018
30-Apr-2018
07-May-2018
14-May-2018
21-May-2018
28-May-2018
04-Jun-2018
11-Jun-2018
18-Jun-2018
Date


Communication Quality Score (1-10)
8
7
6
5
4
3
2
1
0
16-Apr-2018
23-Apr-2018
30-Apr-2018
07-May-2018
14-May-2018
21-May-2018
28-May-2018
04-Jun-2018
11-Jun-2018
18-Jun-2018
Date


Fatigue Score (1-5)
4
3
2
1
16-Apr-2018
23-Apr-2018
30-Apr-2018
07-May-2018
14-May-2018
21-May-2018
28-May-2018
04-Jun-2018
11-Jun-2018
18-Jun-2018
Date


Emotional Intelligence Score (1-5)
5
4
3
2
16-Apr-2018
23-Apr-2018
30-Apr-2018
07-May-2018
14-May-2018
21-May-2018
28-May-2018
04-Jun-2018
11-Jun-2018
18-Jun-2018
Date

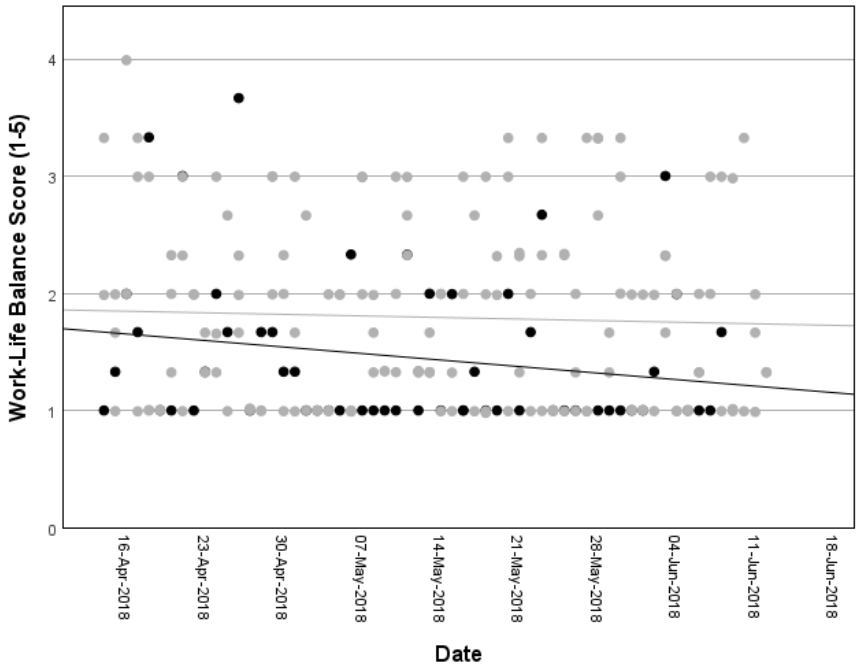
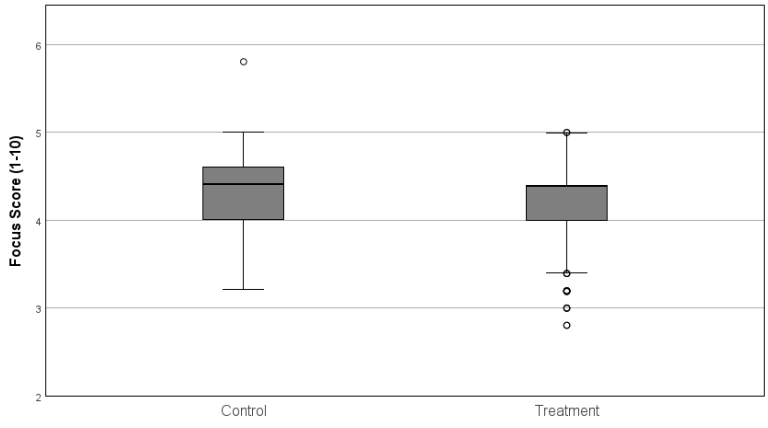
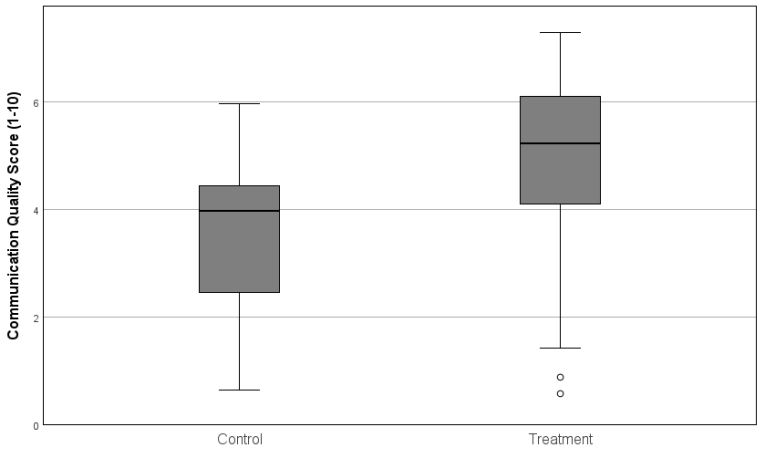
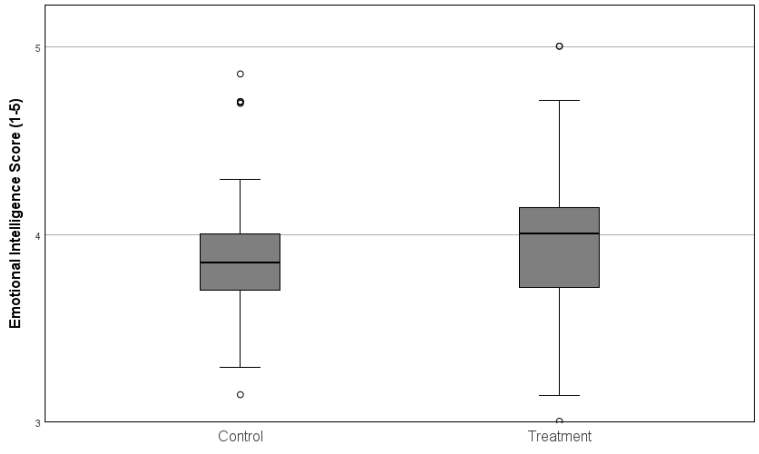
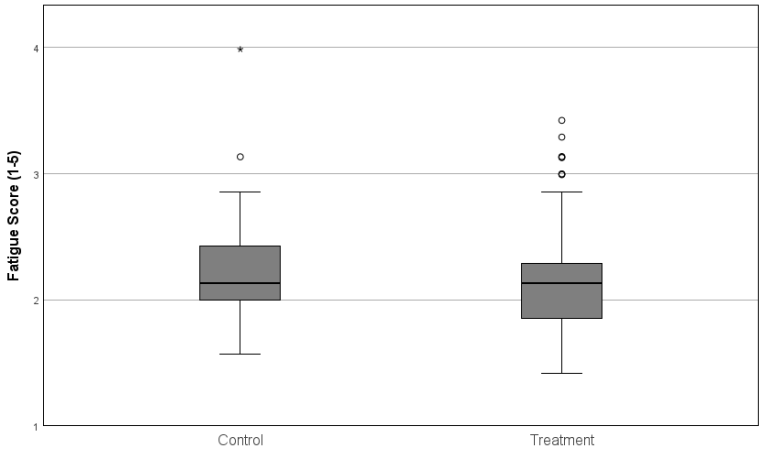


Figure 5.4. Data plots for Survey 2 metrics.




Communication Quality Score (1-10)
0
2
4
6
Control
Treatment


Emotional Intelligence Score (1-5)
3
4
5
Control
Treatment


Fatigue Score (1-5)
1
2
3
4
Control
Treatment

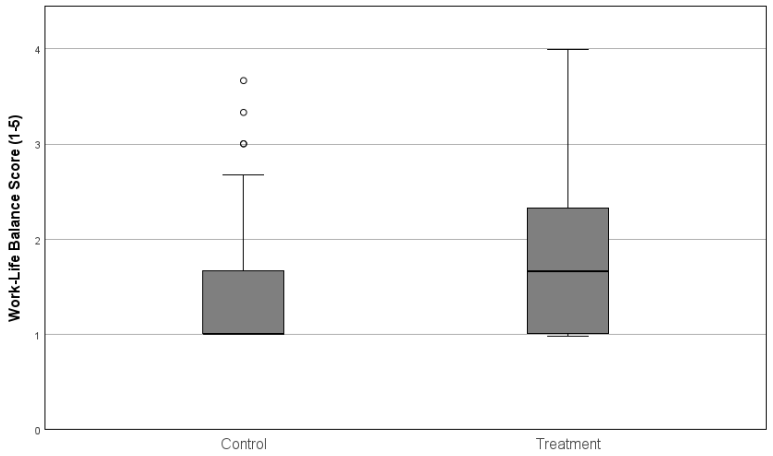


Figure 5.5. Box plots for Survey 2 metrics.

5.5.2.2 Mindfulness biometric results

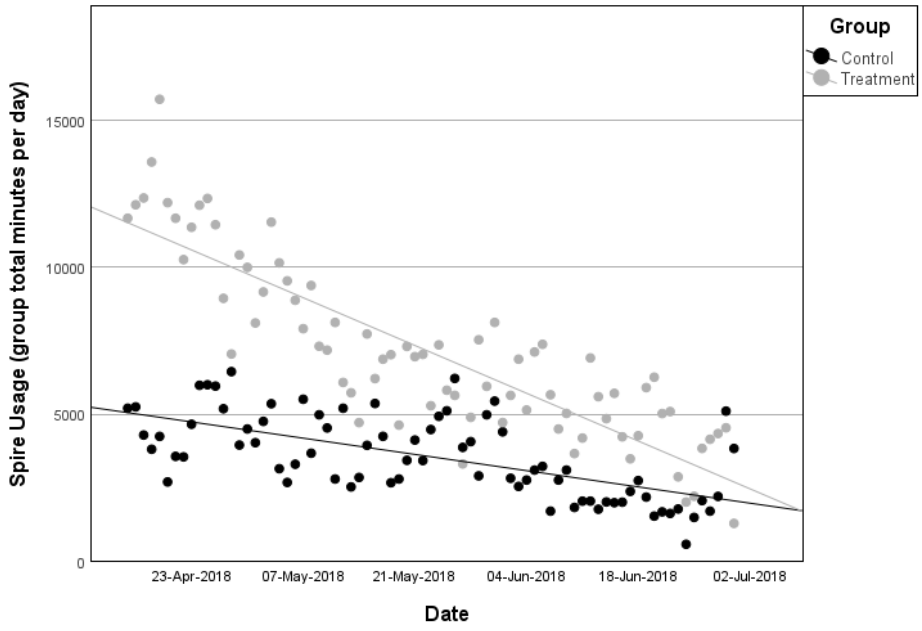
Spire data from all experiment participants also featured non-parametric data and was subject to the same Mann-Whitney between group testing as survey data. There are no outliers to be removed, likely attributable to the non-subjective nature of the data. Spire data also suffered a sharp decline from 15,000 total minutes per day for the treatment group at the start to less than 5,000 minutes per day by the end of the study.

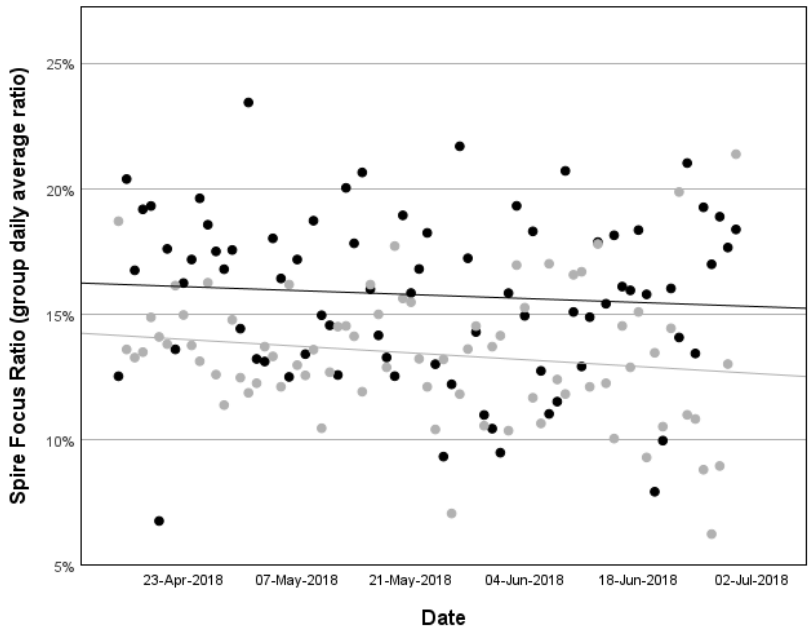
Results in Table 5.7 show significantly higher usage, calm and sedentary scores in the treatment group, as well as significantly lower focus scores. Slope analysis shows that the control group featured significantly different changes over time for usage score (slower decline) as well as activity scores (increasing) and sedentary scores (decreasing). The latter two are likely strongly coupled. The increased scores in calm and decreased focus provides some evidence of training effect, possibly in terms of open-awareness rather than focused attention. Spire data and box plots are provided in Figure 5.6 and Figure 5.7 respectively.

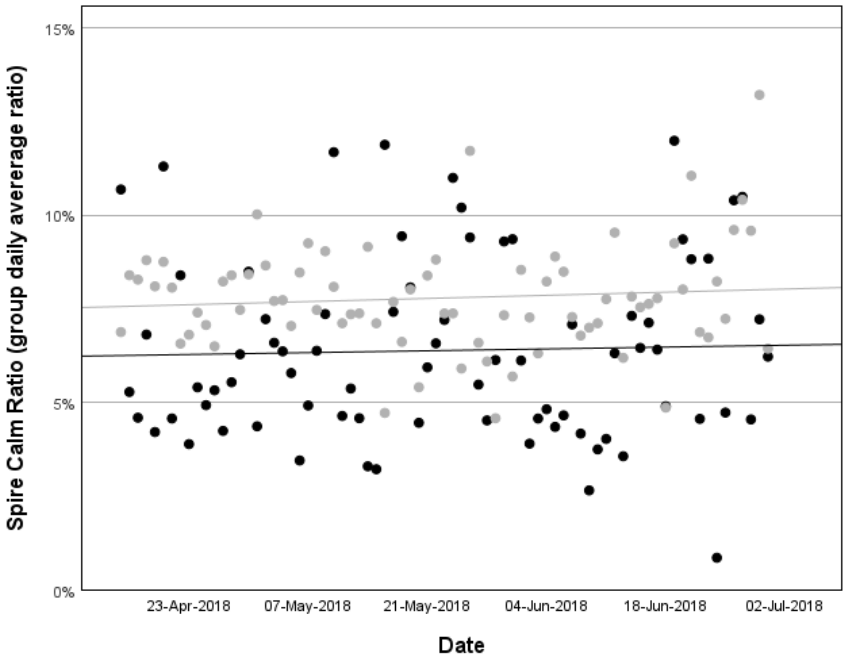
Table 5.7. Spire biometric data between group differences (Mann-Whitney U Test)

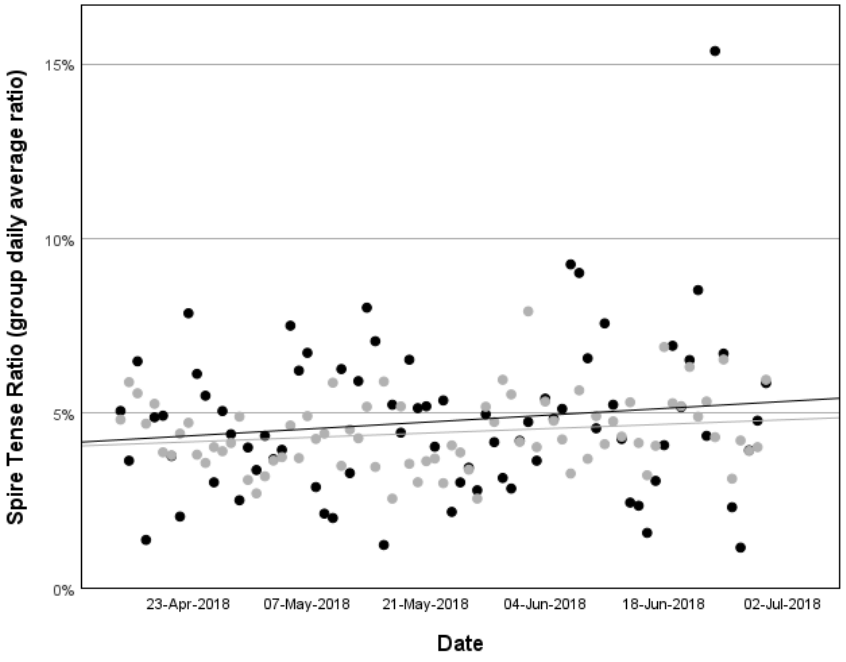
Metric	Control Group mean rank	Treatment Group mean rank	Exact significance p (two-sided)	ANCOVA significance p
Spire Usage	41.61	86.66	<0.001***	<0.001***
Focus Ratio	83.00	56.03	<0.001***	0.379
Calm Ratio	49.05	81.16	<0.001***	0.371
Tense Ratio	69.39	66.10	0.630	0.647
Active Ratio	73.09	63.36	0.153	<0.001***
Sedentary Ratio	57.77	73.70	0.012***	<0.001***

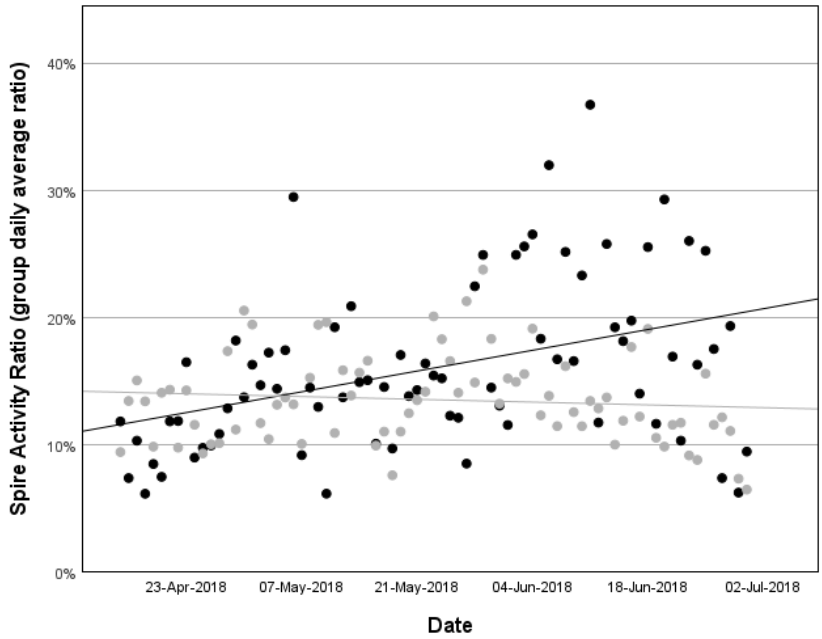
*** $q \leq 0.05$, $p \leq 0.05$, $q > 0.05$**, *$p > 0.05$, *effect size* $r > 0.5$






Spire Calm Ratio (group daily avererage ratio)
15%
10%
5%
0%
23-Apr-2018
07-May-2018
21-May-2018
04-Jun-2018
18-Jun-2018
02-Jul-2018
Date


Spire Tense Ratio (group daily average ratio)
15%
10%
5%
0%
23-Apr-2018
07-May-2018
21-May-2018
04-Jun-2018
18-Jun-2018
02-Jul-2018
Date



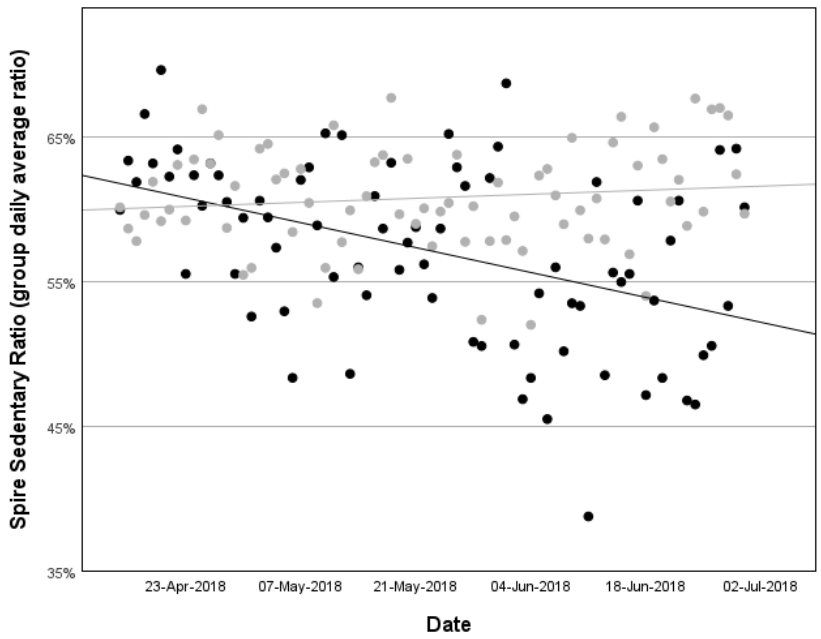
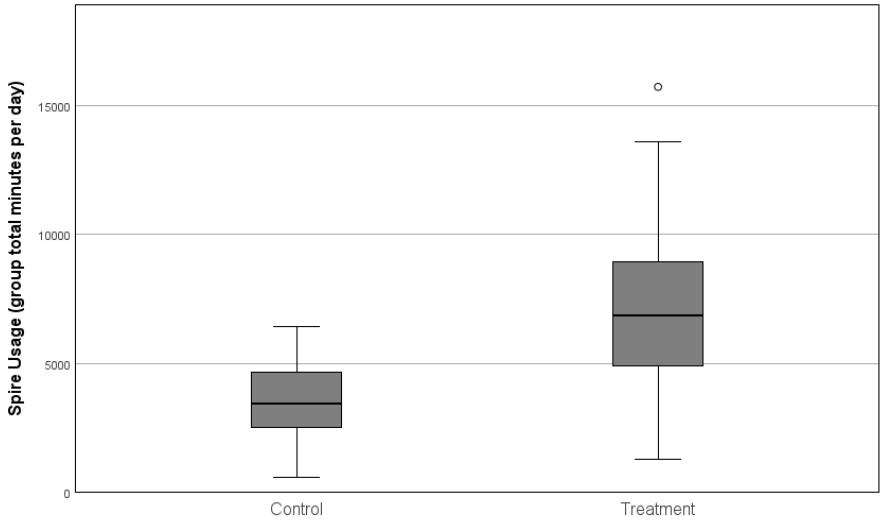
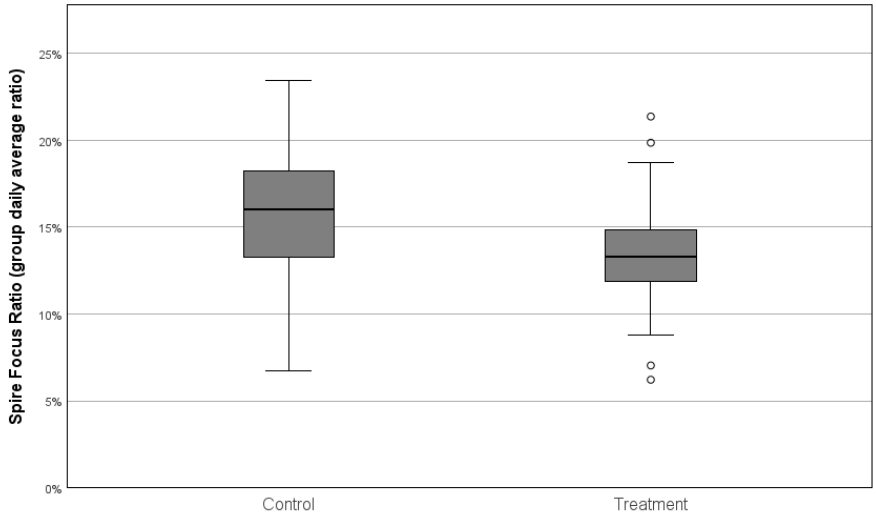
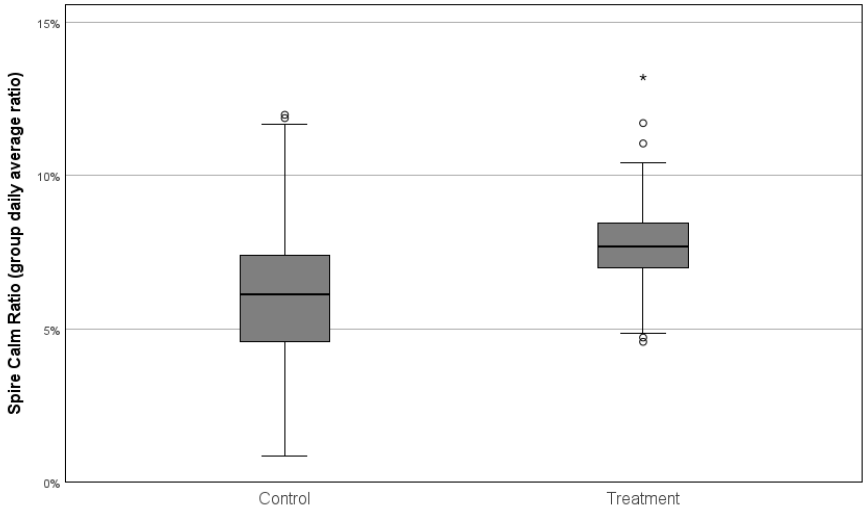
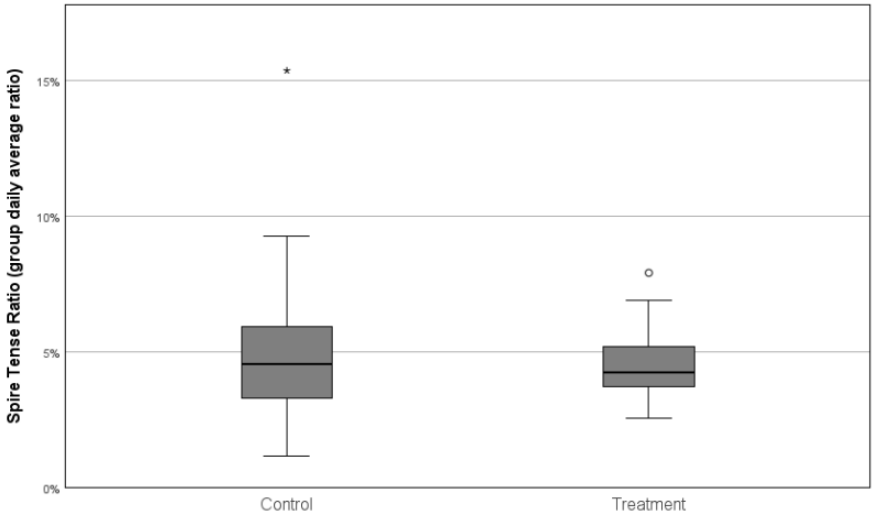


Figure 5.6. Data plots for Spire biometric data.


Spire Usage (group total minutes per day)
15000
10000
5000
0
Control
Treatment


Spire Focus Ratio (group daily average ratio)
25%
20%
15%
10%
5%
0%
Control
Treatment


Spire Calm Ratio (group daily average ratio)
15%
10%
5%
0%
Control
Treatment


Spire Tense Ratio (group daily average ratio)
15%
10%
5%
0%
Control
Treatment

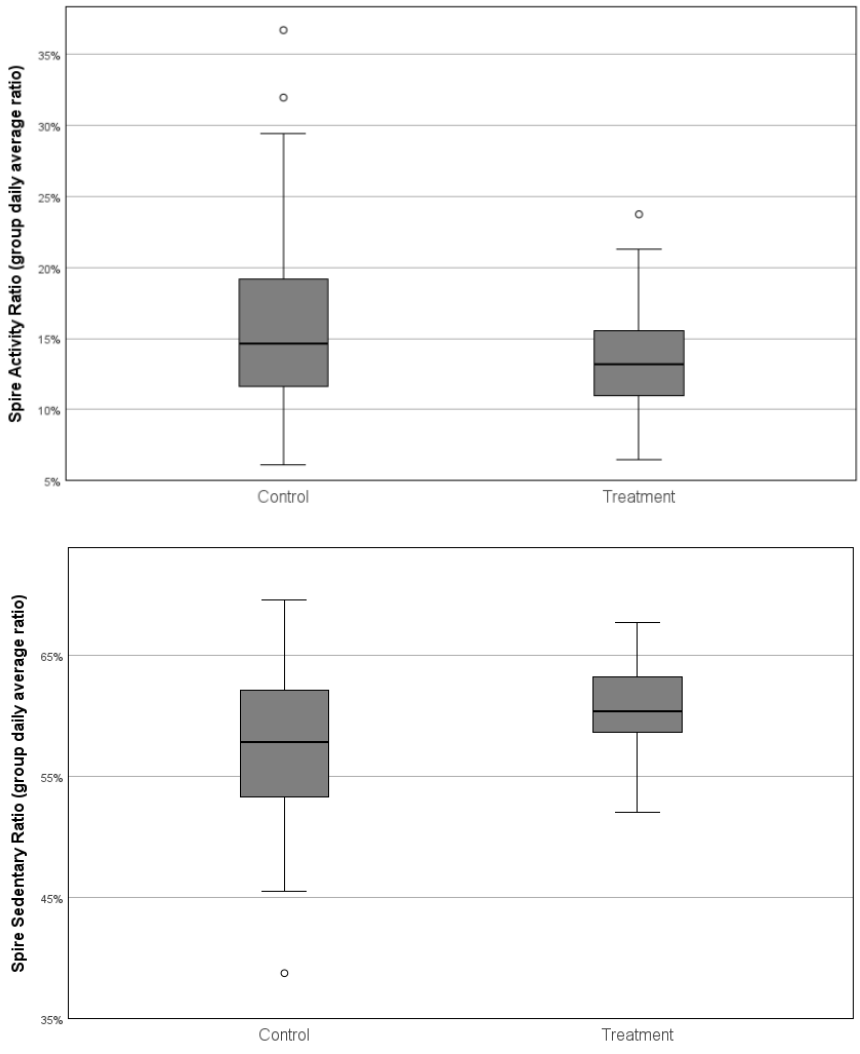


Figure 5.7. Box plot for Spire biometric data.

5.5.2.3 Flight Safety Performance Results

Shapiro-Wilks test for normality indicates that several DFCP (sub)indices are non-parametric. Between-group differences are evaluated with Mann-Whitney U Testing with significance tested at the 0.05 level, BH FDR correction and effect size monitoring. Table 5.8 and Table 5.9 provide the between group comparison results for all DFCP (sub-) indices for flight phase and competencies, respectively. A higher mean rank indicates higher performance. Despite generally hypothesizing that the treatment group will feature higher performance, a two-tailed significance test is still appropriate because it is unknown whether

there may be (unintended) reversal effects where mindfulness treatment may negatively impact a competency or flight phase.

Performance in the holding/startle event (lightning strike) was the only DFCP metric featuring a significant difference between the group (treatment group outperformed the control group), although the test strength was not strong enough to survive FDR *p*-adjustment. Looking across all metrics, in most cases the mean rank performance of the treatment group was higher than the control group, but these differences are not significant despite some effect sizes ("After Landing" phase, and "Problem Solving & Decision Making" and "Startle & Surprise Management" competencies) suggesting that non-significance may be the result of a limited study power with N = 9 for DFCP comparisons. It is also worth observing that the treatment group featured a considerable yet non-significant increase in performance in the "Approach" flight phase which contained the anti-skid failure surprise event. Figure 5.8 and Figure 5.9 show the box plot distributions for both DFCP sets.

Table 5.8. Mann-Whitney U Test between group comparison of DFCP scores per flight phase

DFCP Flight Phase Measures	Control Group mean rank	Treatment Group mean rank	Exact significance p (two-sided)
Total	3.88	5.90	0.286
Paperwork	5.00	5.00	1.000
Preflight Flight deck	5.38	4.70	0.730
Taxi	5.00	5.00	1.000
Takeoff/Climb	5.00	5.00	1.000
Cruise	4.13	5.70	0.413
Descent	3.75	6.00	0.286
Approach	4.38	5.50	0.556
Go Around	5.38	4.70	0.730
Holding/Startle Event	2.88	6.70	0.032** (r = 0.732)
Approach	4.25	5.60	0.556
Landing	4.38	5.50	0.556
After Landing	3.50	6.20	0.190* (r = 0.586)

*** $q \leq 0.05$, $p \leq 0.05$, *q > 0.05***, *p > 0.05, effect size r > 0.5*

Table 5.9. Mann-Whitney U Test between group comparison of DFCP scores per competency

DFCP Competency Measures	Control Group Mean Rank	Treatment Group Mean Rank	Exact-significance p (two-sided)
Application of Procedures	4.75	5.20	1.000
Communication	3.63	6.10	0.190
Aircraft FPM - Automatic	4.38	5.50	0.556
Aircraft FPM - Manual	4.38	5.50	0.556
Leadership and Teamwork	4.88	5.10	1.000
Problem Solving and Decision Making	3.38	6.30	0.111* (r= 0.545)
Situation Awareness	4.25	5.60	0.556
Workload Management	4.13	5.70	0.413
Startle & Surprise Management	3.75	6.00	0.286* (r= 0.563)

*** $q \leq 0.05$, $p \leq 0.05$, $q > 0.05$**, *$p > 0.05$, *effect size* $r > 0.5$*

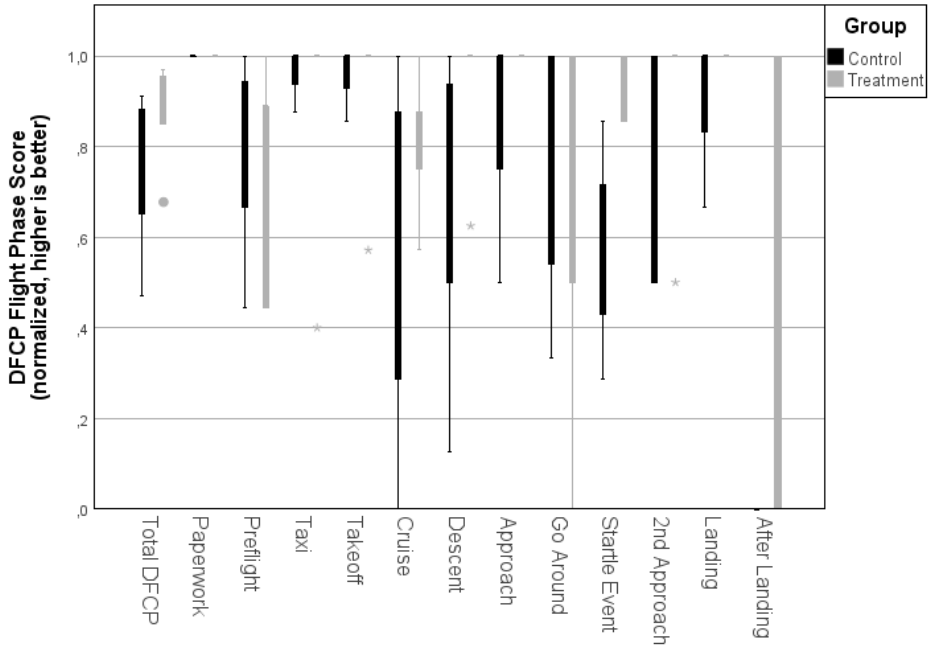


Figure 5.8. Box plot comparison of group DFCP scores per flight phase (in scenario sequence)

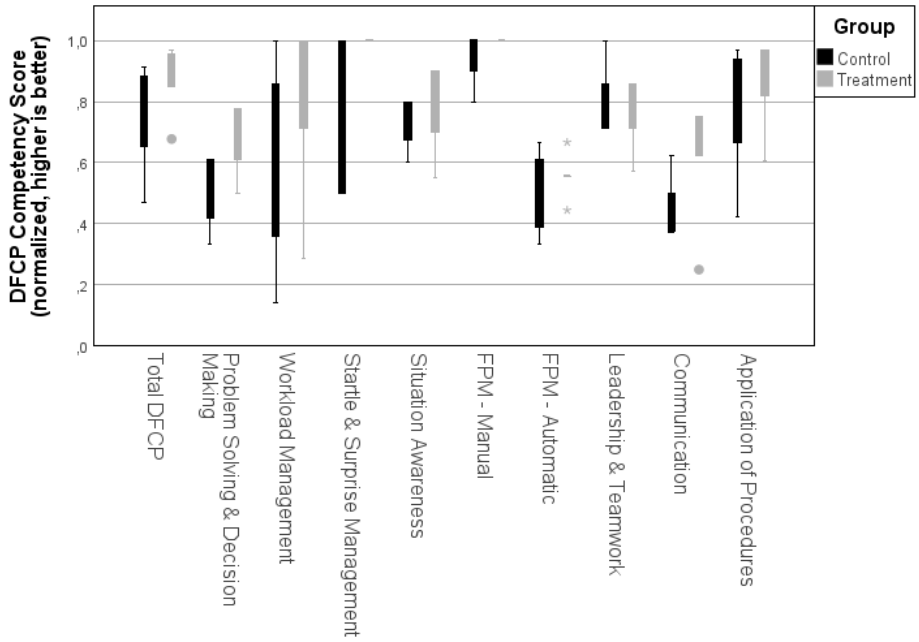


Figure 5.9. Box plot comparison of group DFCP scores per competency (in order of Table 5.4, excluding fatigue management)

5.5.3 Study conclusions

The study provides initial, albeit tentative, evidence that mindfulness training may improve pilot management of opaque situations. Some differences in neurocognitive ability were identified between the study groups by means of data from Survey 2 and biometric Spire data. The treatment group featured significantly higher scores for communication quality, emotional intelligence and calmness, as well as significantly lower focus scores and ratios. Furthermore, biometric data also indicated lower activity ratios and higher sedentary ratios in the treatment group. These results suggest that there may be training effects present.

In terms of competencies, the three "bridging" competencies which were hypothesized to have the strongest treatment effect ("Problem Solving & Decision Making", "Workload Management" and "Startle & Surprise Management") all feature a treatment group showing a higher mean performance than the control group, although not statistically significant. The flight phase DFCP metric for the startle event did show a significant difference (pre-FDR correction), where the treatment group outperformed the control group. A similar, weaker result was found for the approach flight phase with the anti-skid surprise event. Furthermore, the strong trends for "Problem Solving & Decision Making" and "Startle & Surprise Management" competencies where treatment crews outperformed

control crew align with the study hypothesis. These results, while encouraging, cannot confirm that any training effect established also transferred to the flight operational context, and required a more powerful study to confirm beyond the False Discovery Rate (FDR).

The greater calmness of the treatment group may be related to improved open awareness and reduced focused attention, which aligns with the neuroscience of expert holistic sensemaking in the context of ambiguity (which startle and surprise induces). This causal link is tentative but may be reason for a targeted study into such meditation nuances. The increased EI may also have supported emotional regulation to down-regulate the amygdala and limbic arousal in response to startle and surprise and improve crew response during the startle event. However, this causal link between EI, calm and focus with startle management is also tentative, as the study featured several limitations which must be observed before such inference can be confirmed.

The study featured three notable limitations: declining, voluntary participation, lack of within group analysis and limited ability to stratify mindful ability. The progressive decline in participation and voluntary nature of Survey 2 reporting and Spire usage introduces demographic skewing (e.g., particularly engaged or disengaged individuals significantly add or subtract to the data pool). The decline in participation also prevented longitudinal evaluation of Survey 1 which was intended to provide greater resolution in pre- and post-study mindful ability.

The second limitation to perform a between group analysis was in part the results of too many incomplete participations, where only a group-based comparison between mindfulness scores and flight safety performance scores is possible (with significant reservations concerning small sample sizes). Future studies may be improved by means of a within group analysis to establish direct correlation between mindfulness metrics and flight safety performance metrics. Such an approach partially mitigates reduced participation risks and provides a stronger inferential framework around training transfer to flight operations, while still able to evaluate a training effect with between-group analyses.

This leads to the third limitation where individual mindful ability and openness to mindfulness training was not accounted for [26]. An individual crew member's level of mindfulness does not only vary due to exposure to this experiment's mindfulness training but may also be affected by spiritual predisposition, prior meditation experience, personality, motivation (particularly in adult learning contexts) and training methods. Mindfulness training can realize some short-term changes in cognition (Remmers et al., 2015; Berkovich-Ohana et al., 2017), but in many cases neuro-cognitive changes are observed after extended

[26] Note: The groups were balanced in terms of expertise, flight experience, gender and age.

practice (Marchand, 2008; Fabio & Towey, 2018). Studies limited in training capacity and time may feature stronger effects if participants are selected and trained based on their mindful ability and openness (or indeed lack thereof), as this provides a stronger stratification to correlate flight safety with.

Other confounding effects that may have played a role in this study are participant mindset and context (e.g., awareness of it being a study focused on mindfulness), small sample sizes random effects and external factors such as fatigue during some simulator sessions outside of the circadian daytime. As such, the results of this study – although promising – are presented in the context of above limitations and cautions.

5.6 Discussion

The shift to a new model for airmanship to facilitate sensemaking to manage opacity is necessary to mitigate the growing brittleness of the aviation system. However, the requisite cultural change in the aviation industry to enable this shift is not without its challenges. At its core, the aviation industry's culture is rooted deeply in a sense of control and predictability, which has matured and functioned well for over a century. To acknowledge that one might not have that fundamental control in the face of current and future system complexity and accept the opacity and ambiguity it conjures, goes against deep-seated beliefs and existing habits in training, design legacies, oversight practices and judicial structures. There are not many engineers and pilots that are willing to admit they're "not sure": yet being comfortable and feeling competent despite having such doubts is precisely the deeper change advocated here. Schoemaker (2004) underscores the reality that highly complex, highly uncertain situations require us to challenge the conviction (rather: illusion) of total (cognitive) control and appreciate the very real limitations of deterministic strategies rooted in that conviction. In other words, growing a tolerance for ambiguity.

The mindfulness training case study presented in this chapter provides a first-hand account of the challenges in changing this mindset and "heartset". In hindsight, this was due to the study being leveraged in personal vulnerability, and insufficient connection with the study's population to effectively navigate these sensitivities. The neurocognitive mechanisms underpinning Airmanship 2.0 are nearly all possibly quite personal in nature, and future studies in this area are advised to mitigate the risks of disengagement by explicitly managing a safe learning and evaluation environment, while carefully monitoring the perception of such studies. Ironically, Airmanship 2.0 makes the case of *empowering* particularly human abilities, rather than diminishing the role of the pilot. However, this intention must be made clear to participants in the study, to distinguish itself from the hubbub arising from more contested cockpit evolutions such as RCO and HAT.

Another important reflection from this case study pertains to the contextualization of the mindfulness (or neurocognitive) training, particularly when dealing with a "hardened" professional such as commercial pilots in a highly procedural, specialized and deterministic domain. The gap between current flight operations and Airmanship 2.0 practices is considerable, and proper *contextualization* is essential to embed the desired neurocognitive sensemaking behaviors into modern flight training and operations. Techniques such as scenario-based training, peer-to-peer facilitated debriefing and redefined sensemaking and decision-making strategies may be effective to allow pilots to anchor new neurocognitive abilities into their routines and mindsets. Principles of adult learning such as self-motivation, relevance to own work and engagement of one's own knowledge and experience are very effective techniques to maintain interest and motivation (Henley, 2017) despite the possibly confrontational, personal and emotional nature of the training new (meta-)cognitive abilities. Such training also requires a safe environment free of judgment and instructors versed in the neurocognition of sensemaking. Within aviation, such advanced, deeply engaging training contexts are currently rare, although a recent publication (Van Leeuwen, 2024) provides an example of such explorative learning.

Beyond operational research into these mechanisms, this shift to foster a learning function implies changes in how the pilot is perceived by aircraft and system designers, training organizations, regulations and the pilot community itself. Section 1.2 described the fundamental *assumption* that pilots are always in control regardless of the situation they experience. A departure from this assumption may require a change toward the type of persons that are comfortable within this less deterministic context. Pilots who enjoy the predictable, compliant and relatively limited nature of the existing pilot role may not be able to motivate or engage themselves to fulfill the learning function in the proposed new pilot role. Such changes in pilot qualification criteria must balance considerations in pilot drafting and retention to ensure that the pilot workforce size remains stable and also the value of a dual-pilot cockpit as an avenue to build the tacit knowledge required for expert intuition.

Another point of discussion observes the legal implications of this shift to a learning function. Although the premise about pilot total control may approach its limit, it does provide clear judicial demarcations between the accountabilities of aircraft manufacturers, operators and pilots. By acknowledging the opacity and ambiguity (as a result of aircraft and operational complexity) and accepting the need for a learning function within the cockpit JCS, accountability for accidents and incidents may shift between these stakeholders. Are manufacturers responsible for creating opaque systems? Are operators and training organizations responsible for too much flexibility in procedures sensemaking approaches? Are pilots responsible for deviating from (or complying with) procedures? While Airmanship 2.0 may introduce new legal tensions, this redistribution of accountability has already begun

in the aftermath of accidents such as Air France 447, Lion Air 610 and Ethiopian 302, where pilot accountability has made room for manufacturer design accountability (Schwab, 2024; Air Data News, 2025; Garcia, 2025; Charpentreau, 2025; Chokshi, 2025; Reuters, 2026). Such judicial discussions may even become a driving force to challenge the now-defunct assumption about pilot total control.

Despite the above challenges, reaching the next level of aviation safety requires a new approach to safety in response to complexity-related ambiguity and opacity: a stronger human sensemaking function may provide the requisite mitigation. Despite technology having offloaded human pilots of many functions, the human's neurocognitive assets may still be the best source of aviation safety and resilience yet. Future cockpit JCS designs can facilitate the Airmanship 2.0 concept by partially offloading human pilots from the executive and compliant end of their currently bifurcated role. This creates the requisite space for human pilots to do what they do best: explore, learn and adapt. Human brains have evolved for this specific function, how to make use of this unique capability is up to us

6
Discussion and reflections

This final chapter returns to the outset research questions and reflects on their answers based on the dissertation's findings. Discussions explore implications for the further implementation of adaptive, learning and sensemaking functions in the flight deck, particularly in the context of Human-AI Teaming, Reduced Crew Operations and maintaining visibility on opacity-related operational safety and risks.

6.1 Response to research questions

This final chapter returns to the dissertation's research questions with insights from the research performed and presented in the previous four chapters. Section 6.2 discusses the larger implications of this research and provides recommendations in the domains of flight safety, aircrew training, flight operations and aviation research in general.

As a first reflection, the behavioral focus within a practically operational context of this research provided distinct challenges to obtain robust and transparent results. Observing highly variable behavior and inferring underlying cognitive states are both challenging, let alone in free-play simulated contexts which in turn had to be carefully calibrated to achieve and maintain immersion. Furthermore, the relatively small sample sizes and the distributed time and locations of the experiments introduced noise in the results. However, overly rigid qualification standards for results may also induce Type II errors, particularly given the explorative and "fuzzy" nature of such behavioral research. To balance results validity with study sensitivity in the context of explorative research, this dissertation observes less robust trends but reports these with their statistical strength, inviting future studies to further validate weaker tests.

Another important caveat is that the research has focused primarily on evolutions of the pilot model, fully aware that this has been done mostly independent from the rest of the flight deck Joint Cognitive System (JCS). Broader JCS considerations as a point of departure for a more cohesive JCS redesign for Airmanship 2.0 will be discussed.

A final note pertains to the methods employed. In the course of this study and during all simulator research activities, new behavioral analysis techniques were developed and used, tailored to the context of human-machine interactions in complex, highly automated environments. While not immediate objectives of the study, these methodological innovations were turnkey for the study to observe the nuanced and often subtle differences in flight crew behavioral patterns. Section 6.2.2 will address the potential of these methods to improve visibility of aircrew behavior in modern aircraft, particularly in the pursuit of more proactive safety management.

6.1.1 Identifying effective behaviors for opaque situations

The first of the three research questions pertains to a more thorough appreciation of flight crew cognitive awareness and control of modern, highly automated aircraft (i.e., fourth generation aircraft).

Research question 1

What specific aircrew behaviors contribute to or reduce control of highly automated commercial aircraft?

This question has mainly been addressed by the studies in Chapters 2 and 3. The explorative study in Chapter 2 and validation study in Chapter 3 both echo similar patterns in the differentiation between effective and ineffective flight crew responses in opaque situations in complex aircraft. Table Table 6.1 provides a high-level extraction of desired and undesired behaviors. Desired and undesired behaviors are not strictly coupled per row but do show a contrasting trend.

Table 6.1. Overview of desired and undesired behaviors in opaque situations

Desired behaviors	Undesired behaviors
Fail-safe attitude (assume opacity, a level of wariness)	Assuming predictable operations
Independent knowledge (rules of thumb), mental modelling of the actual situation/system	Overreliance on checklists
More time and attention to manage uncertainty and contingencies	Too much short-term, reactive management of the situation
Appreciating procedural limitations	Not using procedures at all
Experience in non-normal ops	Limited uncertainty management, low willingness to learn and make sense
Shared diagnostic approach and task division	Poor collaboration and task sharing
Use automation where it can help	Completely avoid automation

While some of the desired behaviors may seem common sense, both simulator studies showed that this wasn't always the case as they featured both high and low performing crews, despite equal crew qualifications. This distribution may be attributed to the fact that desired behaviors include a level of wariness and openness to (the likelihood of) opacity, and an explicit, conscious valuation of the useability of procedures and automation. Such an attitude conflicts at some level with the safety by reproducibility paradigm still predominant in commercial aviation.

Another key finding is the value of creating time for sensemaking and diagnostics. Crews explicitly and effectively creating time (e.g., through short-term flight planning) created an atmosphere conducive to navigate the opacity they were presented with. In contrast, crews rushing to return to the ground without duly mitigating the situation's threats were considerably less effective. There were indications that this may be coupled with a feeling of discomfort in such situations. Further familiarization and confidence-building in opaque situations may improve crews' ability to make time for sensemaking.

Finally, a lack of diagnostic structure, dismissal of procedures and avoidance of automation entirely were all seen to be performance-reducing. Although an *overreliance* on procedures or automation was not desired, these aids can still play an important role in preventing errors and reducing workload. The nuance is whether flight crew engaged such supporting elements in the JCS as part of their own plan, rather than following them in a dogmatic way. Resolving this adaptation-compliance bifurcation is a key objective of the Airmanship 2.0 concept, which proposes nesting the deterministic within a non-deterministic control paradigm, as described in Chapter 4.

While the above results are obtained from a whole-population analysis in both experiments, the experiment in Chapter 3 also explored the (training) effectiveness of a novel problem-solving strategy and associated Risk Information System (RIS). A between group analysis could not verify that groups with the strategy and/or RIS outperformed an untrained and unsupported baseline group. This may be attributable to the brevity of the training (only several hours of classroom and simulator familiarization before the actual experiment flight). Another explanation may be that making crews consciously aware of their actions and familiarizing with a new sensemaking process may result in them re-navigating certain phases and steps, which was observed in some instances. Furthermore, the small group sizes may have exposed the between-group analysis to random populations effects. An experiment with the same group flying two flights – one before training and one after – may result in a better pairwise training effect study (compensating for learning effects).

6.1.2 Bridging adaptive behaviors with compliant behaviors

The desired behaviors from research question 1 may not simply replace existing flight crew behavioral standards without duly mitigating any negative impact on existing safety mechanisms connected to the pilot model, such as procedural compliance. Chapter 4 addresses this integration by an explicit deconstruction of the existing safety paradigm and proposing a responsible integration of desired adaptive behaviors to manage opacity.

Research question 2

How may an adapted pilot model integrate the desired behaviors with existing safety paradigms in commercial aviation?

Chapter 4 presented a clear case describing the current brittle nature of aviation safety performance related to limited aircrew adaptability. While the current pilot model centered on compliance and safety by reproducibility is part of the existing high safety performance—1.32 accidents per million flights in 2025 according to IATA (2026)—the (unintended) limited adaptability of flight crew is hindering further improvements in safety. Furthermore, future aircraft featuring more complexity and opacity (including perhaps Human-Artificial Intelligence Teaming—HAT) as well as the possibility of reduced crew operations, risk keeping aircrew further "out of the control loop", exacerbating the brittleness problem. Lastly, the diametric opposition of the reproducibility and adaptability paradigms strain the pilot model so long as the industry (implicitly) presumes flight crew proficiency in both modalities. Recent accident typologies show that this is a false presumption, and that there remains a control gap in modern, complex aircraft.

The resolution to this problem lies in mimicking the concept of compound materials. By explicitly nesting compliant behaviors within a framework of sensemaking, such a compound pilot model—Airmanship 2.0—provides a framework in which both compliant and adaptive behaviors are explicitly valued, protected against proficiency decay and engaged in management of both transparent and opaque situations. The following key factors underpin the theoretical rationale of this model:

1. Application of sensemaking becomes the default rather than incidental. This prevents skill decay in sensemaking abilities including self-regulation and adaption of mental models;
2. Compliance and procedures no longer define the entire solution, yet still contribute to reducing slips, lapses and errors and operational efficiency;

3. SOP's and checklists are appreciated together with underlying assumptions and limitations to evaluate their value in further diagnostics or recovery actions; and

4. A base appreciation of opacity in modern aircraft prepares crew for future (possibly even more complex and opaquer) aircraft and systems rather than (inaccurately) assuming full system transparency.

The Airmanship 2.0 model shown in Figure 6.1 connects with the existing pilot competency framework targeted in CBTA and EBT training programs, particularly as non-technical competencies (e.g., leadership, workload management, situation awareness) have been put at equal footing with traditional technical competencies (e.g., procedures, flight path manual control). However, the Airmanship 2.0 model also challenges the competency approach by proposing a nesting of compliance within a primary sensemaking response to release the bifurcation strain between compliant behaviors and adaptive ones. Given the overwhelming exposure to transparent operations for the majority of commercial flights, providing additional structure to maintain proficiency in adaptive responses may be required for such behaviors to engage when they are needed.

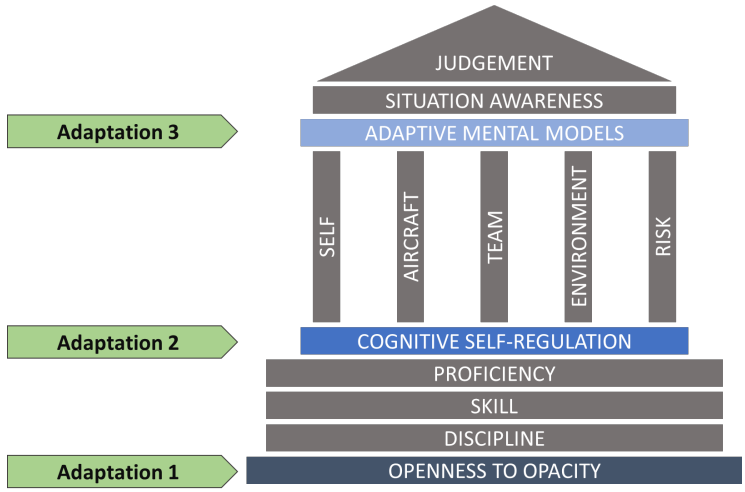


Figure 6.1. Airmanship 2.0 model with three key adaptions to Kern's airmanship model (Kern, 2010)

6.1.3 Training and operationalizing Airmanship 2.0

The restaging of sensemaking as the default response of flight crews required a deeper understanding of aircrew training and operationalization of sensemaking. Chapter 5 approached the sensemaking process at a neurocognitive level to identify how to train the cognition behind sensemaking.

Research question 3

What are the training and operationalization implications of the newly proposed pilot model?

A more detailed breakdown of the cognition of sensemaking identifies three, yet distinct key sensemaking archetypes. Table 6.2 below outlines their applicability as well as training and operationalization considerations.

Table 6.2. Overview of sensemaking archetypes

Sensemaking archetype	Useful for opacity featuring	Cognitive mechanism	Training considerations	Operationalization considerations
Conceptual modelling	Complexity, low ambiguity	Conscious, rational ("cold" executive function)	Stimuli sensitivity training (allo-centricity)	Accessible for relative novices; Can be structured (e.g., T-DODAR/FORDEC); Can be supported with targeted JCS improvements
Heuristic comparison	Low complexity, high ambiguity	Mostly unconscious, affective ("hot" executive function)	Emotional regulation training; explicit learning (training scenarios)	Requires expertise in the form of a heuristic experience base; Connect with analytical approach to improve information cue accuracy for heuristic appraisal;
Holistic appraisal	High complexity, high ambiguity	Unconscious, affective, direct stimuli valuation (a-rational function)	Emotional awareness training; implicit learning (tacit knowledge and symbology)	Also requires expertise, as well as explicit operational latitude for holistic appraisals (e.g., not constricting problem solving and decision-making standards);

Situations featuring neither complexity nor ambiguity feature little opacity, do not require further sensemaking and in general can immediately be managed through pre-determined responses such as checklists and procedures. In terms of embedding Airmanship

2.0, it is important that flight crews are trained to recognize both *complexity* and *ambiguity* opacity dimensions, in order to engage the most appropriate sensemaking mechanism.

The three sensemaking mechanisms are not equally available to all aircrew members. Heuristic and holistic sensemaking function best in the context of ambiguity but also require pre-developed expertise for intuitive cognitive functions to add value to the sensemaking process. As such, both training and operationalization differentiate between novice and expert crew members. For novices however, the focus on conceptual sensemaking requires a down-regulation of internal cognitive "noise"—i.e., affective responses and Default Mode Network (DMN) activity. This down-regulation improves sensitivity to external stimuli, therefore building a more accurately tuned set of experiences as the novice develops into an expert. In this way, differentiating novices in both training and operational sensemaking develops future domain experts which have improved heuristic and holistic responses. In other words, they develop an "experiential database" which has richer and more accurate cues connected to experiences, allowing those experiences to be more readily and accurately recalled in future heuristic and holistic appraisals. Mindfulness training in *focused attention* is a strong candidate to train such novices, together with emotional regulation techniques which have been the focus of multiple startle and surprise recovery studies (Field et al., 2018; Landman et al., 2020; Duchevet et al., 2024; Vlaskamp et al., 2025).

As novices become experts, the focus of neurocognitive training shifts from a down-regulation of internal "noise" to a calibrated receptivity of internal signals. This is important to widen the aperture to deeper cognitive abilities outside of the conscious experience. This may be realized by shifting mindfulness training toward *open awareness,* coupled with simulated training that exposes experts specifically to increasing levels of ambiguity and complexity to practice these neurocognitive sensemaking mechanisms. In this way, these sensemaking mechanisms enable all three key adaptations proposed in the Airmanship 2.0 model (*openness to opacity, cognitive self-regulation* and *adaptive mental models*). Further operationalization must also review JCS developments toward sensemaking-supportive functionalities. Operationalization of Airmanship 2.0 may also include an explicit pre-flight task distribution between novice and expert crew members in the flight deck, depending on the nature of the opacity encountered—i.e., whether the novice or expert is in the sensemaking lead.

A practical experiment evaluating pilots after receiving neurocognitive mindfulness training investigated the potential for such training to improve flight crew responses to opacity. The study was able to provide initial, albeit tentative, evidence that mindfulness training with pilots may contribute to improved management of opaque situations. Trained crews featured higher scores for communication quality, emotional intelligence and calmness,

and lower scores for focus. Outcome performance of trained and untrained crews exhibited some improvements in "startle & surprise management", "problem solving & decision making" and "workload management", which were also hypothesized to be most positively impacted by the training. Strong effect sizes warrant continued investigation of this training effect with a higher stochastic power (i.e., larger population, more scenarios), possibly by means of a meta-analysis.

A final important note on training and operationalizing such sensemaking mechanisms concerns the challenges around pilot acceptance and internalization of such training. The mindfulness study experienced considerable (and somewhat unexpected) resistance and decreased participation related to both skepticism and avoidance of discomfort. The very personal nature of this mindset change, as well as confronting trainees with their (in)tolerance for ambiguity provides specific training challenges to develop an individual's openness to opacity. These results are from the limited scope of a simulator study featuring mindfulness training across two months. More robust results may arise if Airmanship 2.0 principles are instilled much earlier in a pilot's developmental career (i.e., in ab-initio or initial type rating training) and continually reinforced (e.g., through scenario-based training) to develop sufficient openness to opacity.

6.1.4 Overarching research question

The investigations into the three research questions provide the requisite insights to be able to articulate a response to the overarching research question addressing the role of the pilot in current and future flight operations.

Overarching research question

How may an adapted pilot model improve the safety performance of highly automated commercial aircraft?

Responding in the most abbreviated sense, the safety-limiting interplay between current flight crew *brittleness* and system and operational *opacity* in existing and future commercial flight operations may be resolved by explicitly recentering the pilot role and cockpit JCS on a repertoire of sensemaking mechanisms in human cognition—both innate and trainable ones. Unpacking the previous statement distils three important insights gained about the human contribution to (future) flight safety.

The first insight is that the required capacity of the flight deck JCS to navigate the opacity dimensions of complexity and ambiguity may readily be found in several human

neurocognitive abilities in sensemaking. In contrast to seeking a technical solution to response to opacity—very likely exacerbating the Self-Reinforcing Complexity Loop (Klein, 2011)—empowering these abilities in the human pilot may improve JCS adaptability when facing opacity, while also relieving the JCS of systemic brittleness.

The second insight is that enabling such sensemaking capacities requires an explicit and careful merging of adaptive behaviors with compliant behaviors to avoid losing benefits of the existing safety-by-reproducibility paradigm. This can be achieved by defaulting aircrews' response to sensemaking, and nesting compliant behaviors within the sensemaking process, for both diagnostic and recovery actions. By restructuring flight crew responses into such a compound-nested model, skill decay in both adaptive and compliant behaviors may be avoided.

The third insight is that such a restructured flight crew model requires distinct changes in flight deck design, aircrew training and flight operations. In terms of design, supportive flight deck systems may be adapted or developed to elicit and support the wider cognitive spectrum of sensemaking, e.g., through the presentation of risks rather than prescribing solutions as existing ECAM and EICAS systems do. In terms of aircrew training, differentiating sensemaking training between novices and experts, as well as early and continued development of tolerance for ambiguity may be done by means of scenario-based training featuring various *modalities of opacity*, particularly given the relatively low rate of opaque operations given the high reliability of existing aircraft. The concept of *Deliberate Performance* (Fadde & Klein, 2010) may well support flight crew upkeep of sensemaking skills through deliberate pre-flight goal setting and post-flight reflection around sensemaking and decision-making activities [27]. Lastly, operators and Original Equipment Manufacturers (OEM's) may explicitly empower intuitive and heuristic responses through revised aircrew standards related to situation awareness, problem-solving and decision making.

The above insights do not imply a "quick fix" to improving aviation safety. Simply providing an additional training course or an operational memo on airmanship will not transform existing and new pilots toward this new sensemaking capacity. Rather, this dissertation signals a critical review of deeper and long-held presumptions around aircrew standards if the industry is to offset the growth in traffic with a commensurate improvement in safety—particularly in regard to future increases in aircraft and system complexity, and/or a reduction in aircrew. Section 6.2 will reflect on these broader industry implications.

[27] This would be during regular line operations, beyond already established operational training programs such as Line Flying Under Supervision (LIFUS).

6.2 Discussion

The research, experiments, results and conversations related to this dissertation invite a broader discussion about the state and direction of the aviation industry, in particular concerning forthcoming flight deck JCS adaptions such as Human-AI Teaming (HAT) and Reduced Crew Operations (RCO), the suitability of aviation safety monitoring mechanisms and the industry's own capacity for resilience. Each are reflected on in the following subsections.

6.2.1 Evolving the flight deck socio-technical system

Humans and automation in modern, highly automated and complex aircraft do not function in isolation. Any changes to the human's (cognitive) role—such as the Airmanship 2.0 concept— may affect other elements in the Flight Deck Socio-Technical System (FDSTS), and vice versa. A fair amount of research has been established concerning the FDSTS of modern commercial aircraft (Hutchins, 1995; Hollnagel & Woods, 2005; Woods, 2019), even with the recent advent of AI (Naikar et al., 2023; Bjurling et al., 2024). The current state of automation is steadily resulting in a so-called *left-over* FDSTS construct, in which most tasks are automated and "everything else" is *assumed* to be filled up by the flight crew (Bainbridge, 1983; Hollnagel & Woods, 2005; Kirwan, 2025). Such FDSTSs have been designed for efficiency and effectiveness of defined operational tasks, procedures and situations. Their foundation of (cognitive) task analyses underscores this deterministic approach. While these FDSTS's are highly effective in this deterministic domain, they showcase a distinct brittle nature when failing beyond the domain of transparent situations (Manzey et al., 2012; Man4Gen Consortium, 2012; Mohrmann et al., 2015; Woods, 2015; Mohrmann et al., 2017; Hancock et al., 2022; Duvechet et al., 2024; Sarter, 2024).

In contrast, Joint Cognitive Systems (JCS) designs for ambiguous, opaque (non-deterministic) contexts are less explored. The very nature of such opaque contexts prevents their explicit a-priori mitigation though prescribed, designed responses and instead requires a JCS that can facilitate sensemaking, adjusted mental models and adapted responses. JCS studies in recent years have explored concepts such as adaptive automation (Javaux, 2015; Suijkerbuijk, 2017; Hourlier, 2018; Calhoun, 2022; Sarter, 2024), AI-based information systems (Eurocontrol, 2020; Kirwan, 2025) and ground-based support systems (ACROSS Consortium, 2016; Schmidt & Korn, 2017; Myers & Starr, 2021). A particular domain where such adaption is being explored is in the enhanced Minimum Crew Operations (eMCO) and Single Pilot Operations (SPO) domains where JCS artifacts may need to provide scalable support or respond to pilot incapacitation (Schutte, 2015; Stanton et al, 2016; Tokadlı et al., 2021; Harris, 2023; Lopes et al., 2026). However, while such JCS concepts explore innovative artefacts for information provision and decision support, they still primarily operate on the

presumption of conscious, conceptual sensemaking (and possibly even only a procedural response) on the human actor's behalf. By reappraising the pilot model to explicitly value the wider spectrum of sensemaking mechanisms, both flight deck JCS artefacts and dynamics can be adjusted to better support these abilities.

In terms of artefact innovations in the context of Airmanship 2.0, future designs of flight deck systems should be the enhancement and support of human analytical and intuitive cognition, not the replacement of it. Flight deck artefacts can roughly be divided into four types (Parasuraman, 1997):

1. **Information systems** provide raw data with little to no interpretation/modification;
2. **Executive systems** automate certain activities (e.g., autoflight and navigation systems);
3. **Analysis systems** calculate higher-order mission variables (e.g., endurance, top of descent planning, airfields available with appropriate equipment, etc.), and
4. **Autonomous systems** have decision-making authority, with or without notification or ability to intervene (e.g., current flight envelope protection systems).

Flight crew sensemaking can most likely benefit best from improved information systems (1) and analysis systems (3) for better organization of information streams and offloading pilots from tasks requiring mental computational effort. Table 6.3 illustrates this with a simplistic division of JCS tasks.

Table 6.3. An illustration of JCS task distribution suited to opaque situations

Given opacity, *actors* are better at…	**Given opacity, *artefacts* are better at…**
▪ Assimilating independent variables ▪ Heuristic pattern recognition ▪ (Re)-prioritizing information ▪ Adjusting/creating mental models ▪ Coordinating with other actors ▪ Hypothesis formulation (causation)	▪ Providing unbiased information ▪ Higher-echelon information (e.g., related system states, rates, co-occurrences) ▪ Processing large amounts of data (filtering) ▪ Maintaining and retrieving records and logs

As an example, consider the case that a first officer is trying to make sense of a suspected fuel leak (i.e., simple but ambiguous, requiring a revised mental model of the aircraft fuel system). In such a scenario it may be helpful to have flight deck *information systems* that can re-align with the actor's mental model restructuring, internal symbology and sensemaking hypotheses. In this scenario, that may be achieved with the automation's ability to have fuel system information and fuel burn rates on the same screen. Some modern flight decks already feature *display swapping*, but this should go further to allow for greater flexibility in information provision, ideally through applying principles of Ecological Interface Design (Vincenti, 1992; Reitsma, 2025).

Concerning analysis systems, this same first officer may, for example, want to compare the fuel burn of the past hours with historic fuel quantities of comparable flights to discover any significant discrepancies. This may also be coupled with enhanced risk-based analytics (e.g., the Man4Gen RIS or total energy management displays) that can directly connect to Theat and Error Management (TEM) practices. Existing aircraft systems provide limited support for such queries, as they feature either limited (e.g., state) information, or (highly) modified information such as master caution warnings and ECAM/EICAS messages and procedures. As a result, the "cognitive middle ground" is currently left to the flight crew, where such mathematical/analytics/interpretive tasks detract from the crew's ability to engage in information collection and option generation and evaluation. Systems that provide pilots flexibility in comparing and analyzing variables and information at multiple levels may be a better use of the digital capacities, in particular the capacity to query and augment large volumes of data.

The insights into sensemaking also drive a review of the flight deck JCS dynamics *between actors* considering the spectrum of neurocognitive sensemaking mechanisms. The most natural role designation would suggest that the first officer engages as a (relative) "novice" in conceptual sensemaking, while the captain is engaged as a (relative) "expert" in heuristic and holistic sensemaking, as required. This may not always be the case, however. For instance, consider the case of a first officer that is more experienced than the captain (e.g., when the first officer is also engaged as an instructor, test pilot or has more hours on type). Another case would be when a captain has not developed the requisite exposure to complexity and opacity (e.g., due to a late-career switch to a more complex next-generation aircraft). In either case, the dynamics may be shared mutually or be reversed completely. For this reason, the operation allocation of the neuro-cognitive "roles" becomes an explicit part of effective Crew Resource Management (CRM). Deliberately establishing effective cognitive collaboration (e.g., as part of the preflight) allows due engagement of both analytical and intuitive sensemaking mechanisms to maximize the coverage of all types of opaque situations.

Beyond the flight deck crew, other JCS actors such as relief pilots, second officers, ATC and pursers/cabin crew may contribute as effective extensions of sensory information collection for and feedback to the flight crew's sensemaking process. The industry explorations of eMCO, SPO and Reduced Crew Operations (RCO) will likely have an immediate and considerable impact on the flight deck JCS sensemaking capacity. Not only does the reduction to a single crew member in the flight deck potentially remove considerable sensemaking versatility, it also potentially creates an expertise succession gap. A two-pilot flight crew facilitates progressive expertise development and expert workforce continuity, without narrowing JCS sensemaking capacity.

In facilitating sensemaking, artefacts must also be able to vary their complexity and level of information-processing they support. Enabling flight crews to shift in the level of control augmentation (e.g., increasing or decreasing levels of flight control augmentation), information abstraction and system autonomy would enable JCS actor-artefact interaction at both granular and abstract levels as required to make sense of the situation. This is particularly relevant given the advent of AI, which has the potential to introduce more opacity in all four types of artefacts (information, automation, analysis and autonomy). The exploration of HAT may therefore benefit from a JCS perspective, where HAT concepts must be measured against their efficiency and cost-reduction value on the one hand, but also their propensity to increase JCS brittleness on the other.

6.2.2 A shift in aviation safety monitoring

Both the results as well as the methods used in this dissertation also signal that a review of aviation safety monitoring practices may be timely, particularly in the domain of human factors and performance. There are two main points of discussion in this regard:

1. Variations in human *cognitive* behaviors (notably cognitive regulation and sensemaking) are critical factors in modern (and future) commercial aviation and require new investigative techniques (lenses) to monitor these behaviors; and
2. The sparsity of accidents and high diagnostic value of insight into a (live) pilot's cognitive processes requires a shift to pro-active and simulated safety monitoring rather than reactive, accident- and incident-based safety learning.

6.2.2.1 Discussion point 1 - Three new investigative lenses

The first point of discussion argues for innovation in human factors investigative tools: a new generation of a pilot model also requires a new generation of human factors frameworks. Classic human factors frameworks such as NOTECHS (Flin et al., 2003) and HFACS (Shappell & Wiegmann, 2001) have contributed to air safety investigation by expanding the field of view of investigators at the time, however they also suffer from a

fundamental caveat: they are primarily classification systems and do not necessarily provide good insight into underlying cockpit processes, and as such are very limited in their ability to explain—and improve—safety performance related to opacity, complexity and ambiguity. This was already recognized by then-director of the French Bureau d'Enquetes & Analyses (BEA) Jean Paul Troadec stating:

> *"This accident [Air France 447] has also taught us that hypotheses used for safety analyses are not always relevant, that procedures are not always applied and that warnings are not always perceived. Only an improvement in the quality of [investigation] feedback will make it possible to detect any weaknesses in the safety model." (Troadec, 2012, p. 8)*

To provide our safety system visibility on opacity-related risks, three specific human factors forensic lenses are suggested:

1. Competency-based root-cause analysis;
2. Cognitive regulation markers; and
3. Behavioral mapping.

The first lens takes a page from the domain of modern aircrew training. Competency-based training (ICAO, 2013; EASA, 2020) has allowed for a much richer account of human factors including cognitive behavior and experiences to better direct individually tailored training trajectories. While the Airmanship 2.0 concept in this dissertation goes beyond the ICAO competency framework in providing a more explicit structuring of pilot behavior, the competency framework does account for several significant Airmanship 2.0 behaviors (i.e., problem solving and decision-making behaviors, situation awareness behaviors and workload management behaviors). Actively engaging this competency framework in safety investigations together with key practices such as the Observe-Record-Classify-Assess (ORCA) root-cause identification methodology, facilitated debriefing techniques and TEM mapping may greatly improve the resolution of human factors findings.

As the Airmanship 2.0 concept underscores the need for the modern pilot's cognitive self-regulation, the second recommended lens in safety monitoring is focused on assessing whether flight crews are effectively regulating their own cognitive and emotional state, or if their cognitive abilities were deteriorated. This is an important lens for current and future human factors frameworks as it observes a pilot's foundational capacity to demonstrate other competencies. Startle and surprise management research in recent years (Field et al., 2018; Landman et al., 2020; Chen, 2025) have investigated multiple startle recovery techniques and marker schemes that can be used to benchmark crew cognitive

recovery. It is important to stress that these studies underscore the fact that startles and surprises are not preventable, and that the forensic focus must lie on crews' application of effective recovery techniques. More to the point, an openness to opacity *explicitly* acknowledges the inherent propensity for surprise and startle. For this reason, the focus of a human factors framework for cognitive recovery should include both markers for cognitive impact (e.g., reduced communication, confusion, disrupted action flows), as well as include behavioral markers related to recovery (e.g., breathwork, taking distance, checking in with colleagues). Similar to competency-based root cause analysis, evaluation of the existence of a startle or surprise effect is best done with crew experiential interviews and may be difficult to infer from only Cockpit Voice Recorder (CVR) data.

The third proposed lens for human factors safety investigation is behavioral mapping techniques similar to those used in this dissertation. To detect behavioral trends that are beneficial or detrimental to safety, it is important to be able to compare different crews, scenarios and courses of actions in their (subtle) behavioral patterns, rather than only the (performance) outcomes. This would permit comparing complex failure management techniques across airlines, aircraft types, incidents and failure types to identify effective (or concerning) patterns much earlier than if only using subjective accounts. The methods presented in both Chapter 2 and Chapter 3 are examples of codifying complex crew behavior and cognition in relation to performance. This method is particularly valuable when able to contrast positive and negative outcomes of similar situations, such as in a simulated training or research context.

6.2.2.2 Discussion point 2 - A shift in safety monitoring activities

The second point at the beginning of Section 6.2.2 refers to a shift from reactive safety monitoring to proactive safety monitoring. That is not to say that there is no proactive safety monitoring currently in place, but rather to clarify that visibility in the human factors and (cognitive) behaviors studied in this dissertation is far greater when 1) crews are available for (immediate) post-flight interviews and facilitated debriefings, 2) limits of the current JCS performance envelope are sought out (rather than waiting for such rare occurrences in flight operations) and 3) both sub- and super-standard crew performances are observed to monitor safety.

The practical implication of this point is that monitoring the safety of human performance in the flight deck should actively recruit simulated training and research programs to gain visibility on cockpit JCS dynamics, undesired trends and effectiveness of aircrew training. Traditionally, training was often the receiver of safety information to mitigate operational risks with aircrew training accordingly. The reverse now also applies: safety receiving information from training (and research) to receive higher resolution insight

into aircrew risks in modern aircraft operations. Ideally, such training-derived safety data are aggregated on a larger (i.e., global type fleet) scale to identify trends even earlier. This may imply a shift in safety management from individual operators to aircraft OEM's and large training organizations.

In addition to learning from simulated contexts, safety insights gained from positive outcomes (Safety II as termed by Hollnagel (2014)) is an equally powerful driver for safety as it increases the volume of information to observe as well as provides examples of *desirable* flight crew behavior, rather than waiting to learn from (increasingly rare) negative outcomes (Safety I). Furthermore, investigating "uneventful" flights under particularly challenges circumstances (e.g., with several threats and potential for error) may be more valuable as it provides clear evidence of what behaviors actually *increased* safety. A practical implementation of such Safety II learning was explored by Dijkstra (2023), in which the "informal bar conversation" of pilots is collected and shared within the company or community for collective learning. In contrast to traditional abstract safety reports and statistics, vivid experiential accounts from another person's experience connect directly with one's own tacit and heuristic knowledge base and symbology, allowing the other's experience to be incorporated as one's own.

This value of stories is also underscored by Klein (2011) who poses that stories provide richer context and clearer situational markers from which better hypotheses stem when analyzing data in opaque, ambiguous situations (i.e., "connecting the dots", as Klein states). The approach by Dijkstra (2023) is effective because it allows direct peer-to-peer learning in both negative and positive experiences. From a Safety II perspective, a story about prevention or prompt recovery may be more valuable than stories about negative outcomes, as it teaches both *recognition* of opacity, the *sensemaking experience* and *outcome* or resulting solution, growing positive heuristic repertoires of presenters and listeners alike. Such Safety II shared learning dovetails well with the practice of *deliberate performance* discussed previously.

The above subsection does not imply that existing human factors investigation frameworks should be doubted at every corner, but rather to extend and evolve them to embrace a more proactive modality. By borrowing new lenses from research and training to increase forensic sensitivity to human cognitive experiences under opaque conditions, safety monitoring programs and investigators can be powerful canaries in the proverbial coalmine, identifying deficiencies in our FDSTS before they manifest as an actual loss of safety.

6.2.3 Resilience of the aviation industry

This third and final discussion concerns the transitional pathway from existing aircrew, flight deck and safety paradigms to the new paradigm and concepts presented in this dissertation. To what extent can a new paradigm with human sensemaking at the center be

achieved through derivative evolution from the current industry standards, or to what extent is a more disruptive transition at hand?

Lessons may be learned from historic social and industrial evolutionary cycles. The current safety performance of the aviation industry is high, yet brittle. This is the result of decades of standardization, compliance and safety by reproducibility is in many ways at the limit case of this safety paradigm, as clearly visualized in Figure 6.2 (adapted from Chapter 4).

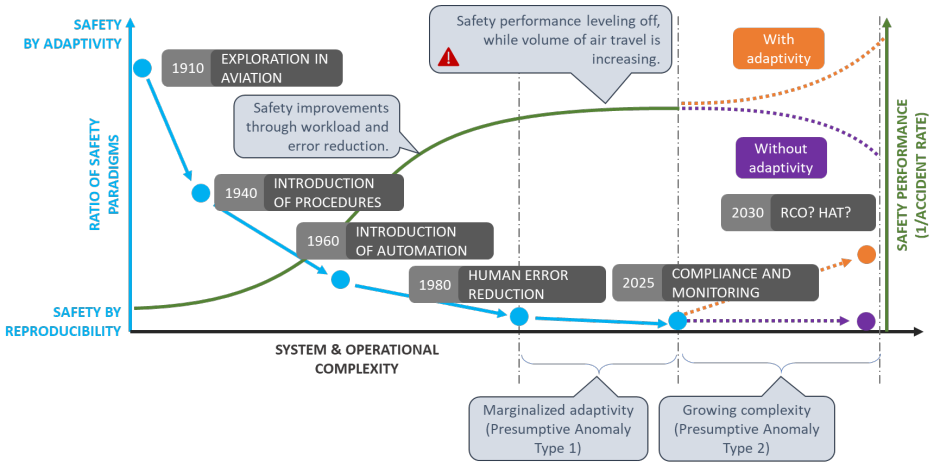


Figure 6.2. Conceptual visualization of the control paradigm ratio (blue) and safety performance (green) over time with future scenarios (orange and purple)

This development is described as a concept of *collective ossification* (Van der Molen, 1989). Van der Molen explains that nearly all social structures and systems (such as our aviation industry) are subject to ossification due to natural social dynamics between an *adaptive majority* and *innovative minority* (Kirton, 1976), progressively strengthening dominant norms and standards, promoting compliance and experience behavioral, intellectual, physical and even genetic consolidation (as applicable). While this process is gradual as illustrated in Figure 6.3, the resulting ossification hampers the adaption of the dominant system to a changing environment (in the case of aviation, the context of increasing complexity and opacity). In many cases, other better performing systems disrupt and replace the status quo, not unlike the *presumptive anomaly* described by Constant (1973) and Vicenti (1990).

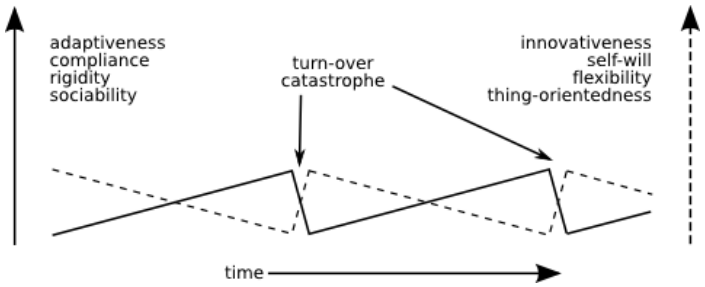


Figure 6.3. Illustration of a turn-over catastrophe after a period of ossification (Adapted from Van der Molen, 1990)

As the cracks of brittleness are showing in the paradigm of compliance and reproducibility, yesterday's design assumptions no longer seem to apply in the new context of complexity and opacity. To avoid persevering to a "turn-over catastrophe", the aviation industry must adopt a resilient, adaptive attitude to this changing environment.

This puts the aviation industry in a similar position as the pilot facing an opaque situation: how to proceed when existing solutions don't work? A process of sensemaking may guide this transition. This begins with challenging our industry's core (implicit) presumptions about pilot total control and proficiency, aircraft system transparency and safety accountabilities that are deeply engrained in design, certification, training and operational regulations and economics. The reality is articulated punctually in an anecdote borrowed from a (humble) pilot colleague:

> *"As a pilot I thought I knew how aileron systems work in a generation four jet, but the simple fact is there are several PhDs being performed on aileron control systems alone, so I cannot possibly know all the intricacies of that system."*

While accepting the limits of such beliefs and departing from the presumption of determinism may be unsettling, accepting *opacity as an expected phenomenon* in engineering, certification and operations allows for a better response than presuming it does not exist. When we recognize this as such, the step to innovate is no longer one of great courage, but one of natural responsibility:

This awareness allows subsequent sensemaking at the industry level. In a way, this dissertation has been just that: *making sense of sensemaking*. By better understanding how flight crews (should) engage with opacity, it provides insight into an emerging operational risk that requires solutions outside of the industry's existing response repertoire. Other research activities, for example in startle and surprise recovery and safety studies like the IATA EBT Data Report (IATA, 2014b), also contribute to industry-level sensemaking all the same. However, the greatest contribution to industry sensemaking lies in evolving our safety

monitoring processes as described in Section 6.2.2, opening the aperture to higher resolution and proactive safety learning from simulated contexts and positive ("normal") operational experiences. In this way, safety programs identify purported "black swans" before they appear as such, materializing the notion of serendipity and guarding against ossification.

The insight of human pilot sensemaking as an effective response to opacity in flight operations requires coordinated changes across multiple industry stakeholders in order to take effect: from design through certification to operation. As a point of departure, Table 6.4 illustrates several considerations for different stakeholders that may contribute to a future-proof flight deck JCS.

Table 6.4. Considerations for industry stakeholders in light of expected opacity

	By accepting the new socio-technical paradigm will feature opacity…	**Considerations**
Engineering…	recognizes the non-determinist nature of modern systems.	Explicate all design assumptions, particularly around system opacity and human operator performance. Devil's advocate to challenge assumptions.
Certification…	regards opacity as a very possible phenomenon in complex systems and operations.	Flight deck and other human-machine interactions must be evaluated for the risk of opacity, and associated mitigations. Performance-based oversight replaces assumed performance by compliance.
Training…	should actively prepare aircrew [28] for opacity management.	Select for and develop a tolerance for ambiguity, cognitive self-recovery and sensemaking techniques based on level of expertise.
Operations…	must accommodate human operators to engage in adaptive behaviors.	Redefine aircrew performance standards and structure procedures and SOP's to actively empower sensemaking.

[28] This example highlights opacity in the context of the flight deck and aircrew but may equally apply to other domains such as maintenance and air traffic control.

Safety investigation…	must gain visibility on opacity and related risks.	Introduce new (human factors) lenses to monitor the risks related to opacity and how human operators manage it.
Accountability…	may shift from the flight crew to training and design as sources of opacity risk.	OEM's may need to assume greater accountability for their designs, and mitigate accordingly (e.g., with training).
Research…	should focus on improving responses to opacity.	Better mapping of interactions between humans and systems, with higher resolution in behavior models.

The above changes may establish resilient behavior at an industrial level, mitigating opacity with an effective pivot to unload derivative strain, rather than a (catastrophic) *turn-over*. This is not only achieved through redefining airmanship, but also redefining *engineeringship*, *inspectorship*, *operatorship* and *investigatorship*. As we design, build, certify and fly with a new kaleidoscope of presumptions (i.e., the expectance of opacity) we may do so fully aware of the risk of ossification, rotating this kaleidoscope every now and then to put our systems in a different light. In other words, allowing us to identify Airmanship *3.0* before we reach the limits of the Airmanship 2.0 model. Only through such a persistent self-reflecting, assumption-challenging capacity can we remain sensitive to the true design limitations, and pivot to new systems before the old ones exceed their due-by date.

6.3 Recommendations

Based on and in extension of the results of this dissertation, the following seven recommendations are made in various domains as indicated.

1. **Research:** Explore and validate neurocognitive sensemaking training for pilots, cross-referencing performance with brain activity monitoring and experiential feedback.
2. **Research:** Explore and validate the effectiveness of flight deck JCS innovations that specifically support sensemaking mechanisms (notably conceptual sensemaking). This may include careful application of AI to support mental model building, taking care to avoid further complexity.
3. **Training:** Define and test training programs that develop the three sensemaking neurocognitive mechanisms (analysis, heuristics, holistic). This may include cognitive regulation techniques, focused-attention and open-awareness mindfulness techniques.
4. **Training:** Develop scenario-based training programs (i.e., through CBTA and EBT frameworks) that specifically challenge and develop the three sensemaking mechanisms (analysis, heuristics, holistic).
5. **Operations:** Operationalize the Airmanship 2.0 fail-safe principle into new problem solving and decision-making standards. This should reflect nesting compliant behaviors within an overarching sensemaking response.
6. **Flight safety:** Establish new human factors forensic techniques to improve visibility on sensemaking and cognition in modern flight decks, also learning from positive experiential accounts.
7. **Flight safety:** Shift monitoring of flight safety from operational monitoring to simulated (e.g., training) environments to access more nuanced dynamics within the FDSTS, as well as test the FDSTS under opaque conditions which are (by the nature of reliability) rare in operations.

6.4 Closing remarks

As we evolve and shape our society, systems and technology, we have incredible latitude to determine how we regard and apply ourselves and others. While our human nature may at times be the cause of our misery and liabilities, we are equally also our liberation in trying times. Seeing ourselves as liabilities as Taylor did, brought us good performance at the cost of great performance.

As globalization, commoditization and emission scrutiny of air travel provides constant pressure to reduce costs, increase performance and increase capacity, the challenge for flight crews is shifting from ensuring reproducibility and compliance on a daily basis, to providing a generative, learning capacity in moments that are few and far between. While aircrew training, flight deck design and oversight all have an important role to play in this shift, it also beckons a deeper reconnection with our human nature and capacity (courage even) to tolerate ambiguity through a willingness to observe and learn, rather than an ossified dependance on determinism. Beyond competence, this draws from individual character, self-belief and self-improvement.

It is precisely those learning capacities that defines our humanity - creativity, understanding, projecting, collaborating – that allow us to strive for a higher standard of performance and resilience in an increasingly dynamic world. This is particularly true in this modern age featuring greater opacity, whether in modern and future flight decks, disruptions in defense, globalizing business or the expansion of technology in our everyday lives.

Enabling and refining these humanities to be resilient was recognized by the Greek and ancient stoic philosophy as *Arete (virtue)*, challenging our wisdom, courage, justice and temperance to become better versions of ourselves. These themes of self-actualization align strongly with Airmanship 2.0 and are the path towards great performance under uncertain conditions. To flourish in such a time, it is upon us to shift our focus away from human error—and embrace the arete of the human asset.

Volare humanum est

Acknowledgements

It takes a village to raise a PhD candidate, and my academic journey is no exception. Where I may have the privilege to present this contribution to the body of knowledge, I would be remiss to not articulate my gratitude, respect and deep appreciation for the academic village I was raised in.

To pursue such an academic ambition alongside a relationship and career has only been possible by the absolute grace of my wife Marjolijn, whose love, support, motivation and patience provided the space in our garden for this thesis to grow roots and bloom. Your ever-present willingness to support me (with an appropriate *push* at times) is more than I may ever have permitted myself to ask of you. I could not have asked for a better wingman to fly through these past years, where you have allowed me to continue to grow alongside you – thank you for being with me on this journey. Alongside my other wing, I thank my parents Bastiaan and Frederike, and my siblings Hans, Emmelie and Aurelia, their dear partners and my extended family for our close family bond, inspired upbringing and at times animated debates - which most certainly have given me an edge in defending my ideas and ambitions.

I have been blessed to have been raised in a supportive academic village stretching across the domains of aerospace and operations, safety engineering, human factors and psychology and training and simulation. It was during my master's education that – in a period of orientation - I met John who not only exposed me to a rich domain of forensic and safety science in aviation but also inspired me to engage in an apply myself to a domain which above all others I felt a moral magnetism to pursue. Where I learned that disagreement was not only required but expected, if we are to move the needle on safety. Where engineering prowess meant reflecting, learning and understand, rather than inventing without regard for consequence. John was the proverbial ignition to my academic rocket, for which I am ever grateful. Although I'm sure he'll find a way to disagree, and I expect nothing less.

A considerable portion of my academic upbringing was in the NLR village, which has allowed this rocket to gain new heights through the support, guidance, teamwork and warm collegialism of many fellow researchers including Joris, Arjan, Edzard, Jeroen, Jenny, Chris, Mick, Petra, Anneke, Harrie and Michel and many, many other close colleagues with which I shared many academic adventures, late nights (and weekends in the simulator), head-cracking brainstorms and high-fives when it all worked out. I also extend my gratitude to the broader academic network I have had the privilege to work alongside with: from Man4Gen and other research consortium teams, to the safety-seeking airlines, innovative organizations and passionate individuals who have been part of this journey, including my fellow PhD

researchers Arthur, Frans and Geert, who inspire and invigorate me to regard the bigger application and relevance of human resilience – both inside and outside of aviation. I extend my particular gratitude to the men and women pilots who volunteered for these and other research studies, and who continue to volunteer their time, professional skills and expertise to make aviation safer for the next generation.

A rocket needs a guidance system, and I am most obliged to my promotors Max and Sidney who volunteered to guide me with their time, expertise and reflective supervision, sharpening my academic steel to a whole new level. Their very complementary perspectives, ruthless reviews and most gracious coaching provided precisely the right conditions for this thesis to find its trajectory and mature. Where this thesis experienced the occasional patch of turbulence from my career and life events, their supervision adapted to accommodate (including some very dear life lessons). Where I accelerated and pushed forward, they directed my energy toward those parts of the deliverance that I had passed in my enthusiasm. I am incredibly grateful for their willingness to take me under their academic wing, creating a special space in the TU Delft aerospace engineering village, and provide the patient preceptorship throughout my less-than-usual PhD journey.

In preparation for landing in the final year of my thesis, I've experienced an immensely close support network of friends who have given me the space and proverbial parachute to grease the landing. Friends who have been by my side as the pressure rose, who helped me find my feet at times, and run, sit, puzzle, cook, fly, chop wood, reflect on life or just being present together at other times. A heartfelt thanks to Alex, who has been immeasurably flexible and patient in our new professional adventure together, provided me the space to complete this thesis in tandem. In the touchdown zone, I extend my thanks to each member of the graduating commission for the dedication, time and interest in the result of over a decade of research, and my gratitude to my paranymphs Nitant and Tim who stand by my side as I bring this chapter of my life to a closure (and make me look sharp doing so).

As I draw these acknowledgements to a close, I cherish fond memories of Fons Claerbout and Alan Miller, two of the most dedicated individuals whom I have the privilege to share both a friendship with, as well as the vision for the future aviator. Both were exceptional airmen, and have their fingerprints etched deeply in this thesis. Both flew to their own clouds too soon. But above all, I hold in my memory how - in their own ways - they exemplified learning and self-reflection, with a humbleness to opacity we may only admire - living Airmanship 2.0 *avant la lettre*.

Appendix A

Man4Gen Quick Reference Card

(For illustration purposes only: layout and font are adjusted to fit in this appendix)

COMPLEX SITUATION MANAGEMENT GUIDE

STABILIZE FLIGHT PATH

- **FLY THE AIRCRAFT**
- **CONFIRM FLIGHT PATH CONTROL (OR IF SEMI-STABLE)**
- **CONSIDER USE OF AUTOFLIGHT**
- **ASSIGN PF/PM**

NO PAGES

IMMEDIATE THREATS

- **IDENTIFY IMMEDIATE THREATS**
- **PRIORITIZE THREATS**
- **THREAT MEMORY ITEMS**
- **CONFIRM THREAT STATUS**

NO PAGES

SHORT TERM PLAN

- **IF POSSIBLE, MAINTAIN FLIGHT AND BUY TIME**
- **CONSIDER FLIGHT PLAN OPTIONS**
  Original destination - alternate destination – holding – land ASAP
- **CONFIRM SHORT TERM FLIGHT PLAN**
- **CONSIDER NOTIFYING ATC/CABIN/COMPANY**

RISK LEVEL PAGE (0)

FLIGHT PHASE RISK PAGE

IDENTIFY SITUATION

- **ACKNOWLEDGE CERTAINTIES, UNCERTAINTIES AND CONCERNS**
  Issues to consider: fuel or time limit? Structure integrity? Controllability/performance? Information reliability? Secondary failures?
  External complication factors such as weather, traffic and routing/destination?
- **CROSS-CHECK SUSPECTED SITUATION**
- **IF SEVERAL POSSIBILITIES: PREPARE FOR WORST CASE**

CATEGORY RISK PAGES (1-4)

PERFORM APPROPRIATE ACTIONS

- **ASSURE THAT ACTIONS WILL BE SAFE, EFFECTIVE AND INFORMATIVE**
- **PERFORM PROCEDURES & OTHER ACTIONS**
- **VERIFY EFFECT OR KNOWLEDGE GAINED FROM ACTIONS**

CATEGORY
RISK PAGES (1-4)

LONG TERM PLAN

- **IDENTIFY EFFECTS OF THE SITUATION**
  Issues to consider: Time/endurance limits? Struct./system functionality? Controllability? Performance? Info. reliability?
- **ADAPT MONITORING TO DETECT IMPORTANT CHANGES**
- **CONSIDER REGULAR FLIGHT PLANNING ASPECTS**
  Weather, company, pax, deferred tasks
- **CONFIRM LONG TERM PLAN**
- **CONSIDER NOTIFYING ATC/CABIN/COMPANY**
-

RISK LEVEL PAGE (0)

FLIGHT PHASE RISK PAGE (5)

Appendix B

Man4Gen Risk Information System (Adapted from Buch et al., 2017)

All risk information pages have been designed to mimic the visual layouts and screen sizes of the Airbus A320 ECAM system pages. Figure B.1 (left image) shows a risk calculation for various flight phases. Calculations should take into account many different aspects of the aircraft state including current altitude/airspeed, location, malfunctions, endurance and other factors such as weather and geographic risks. For the sake of an experiment, the pages used in the study presented a select number of "drawn" pages that would change based on the scenario time and if particular (failure) events occurred. In this particular page, it is clear that the Landing/Go-Around flight phase presents the most risk and includes two pieces of information that have the greatest contribution to this risk ("Reduced Climb Rate" and "Yaw Oscillations"). This should drive crews to mitigate with appropriate threat and error management briefings and strategies.

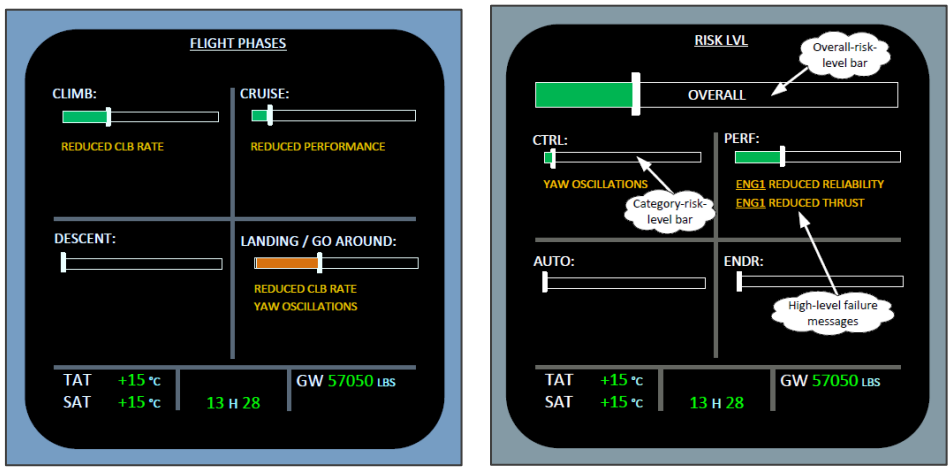


Figure B.1. RIS flight phase risk page and risk overview page

In the risk overview page (Figure B.1, right image) the RIS provides crews with another "cross-section" of risks, namely those associated to different aircraft state parameters: controllability, performance, automation and endurance. In this example, *aircraft performance* is most affected, with two failures driving this risk assessment ("ENG1 reduced

reliability" and "reduced thrust"). This page also provides an "overall" score, which supports the other categories in driving crew awareness to mitigate and brief accordingly.

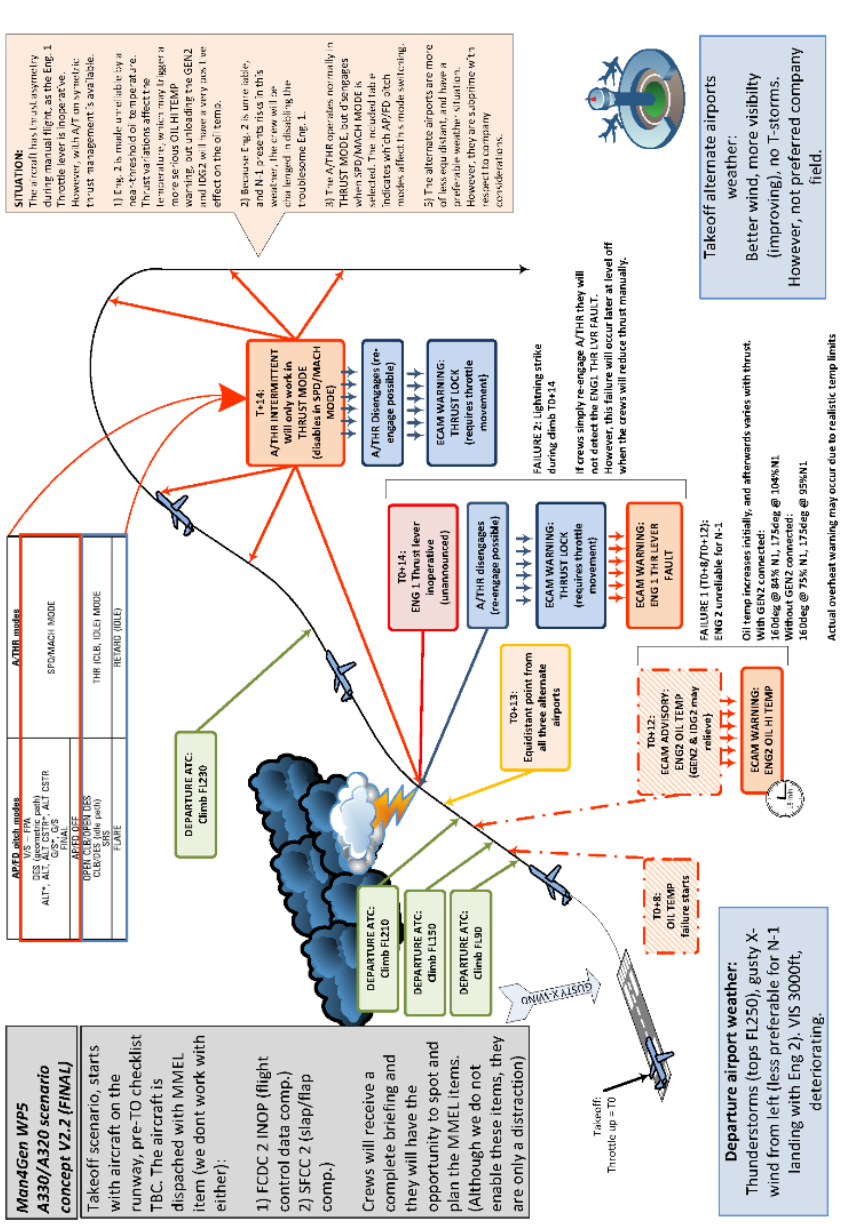


Figure B.2. RIS detail page and picture of the RIS in the AVES simulator

Each risk (in this case *aircraft performance*) can be selected to inspect more details about the systems considerations and how particular failures are driving its risk evaluation. These detailed pages also include *suggested* actions or considerations to support the crew's threat and error management. However, the verbatim is carefully selected as to not stimulate automation bias and prescription-based resolutions. Figure B.2 illustrates how the RIS is embedded in the natural cockpit environment, in order to minimize the impact on crew training and make use of human-machine interactions already familiar to the crews (e.g., ECAM system page selection).

Appendix C

Chapter 3 scenario schematic



Appendix D

Chapter 3 scenario DFCP items

DFCP Item	ENG2 managem't	ENG1 managem't	Route/ situation managem't	Comm./ Interaction	A/THR managem't	Other flight tasks
Avoiding storm cells in path			X			
Event: OT Advisory						
Verbalize oil temp rising	X			X		
Refer to QRH ADV page	X					
Reduce Thrust	X					
GEN2 OFF	X					
IDG 2 OFF	X					
Collect information of weather at nearby fields			X			
Discuss weather situation before diversion decision. home base deteriorating (& enroute to diversion options).			X	X		
Do not land at home base			X			
Event: OT ECAM Warning						
ENG2 IDLE before shutting down	X					
APU GEN ON	X	X				
APU ON Before ENG2 shutdown	X	X				
Use ENG 2 within oil temp limits	X					
N-1 descent and approach planned			X			
Identify ENG1 throttle issue & effects		X				
Event: THR LVR FAULT & A/THR FAILURE						
Own initiative to find FCOM		X				
Call maintenance/FT/Dispatch own initiative				X		
Declare MAYDAY/PANPAN to ATC				X		
Identify THRUST LOCK		X				
Keep/restart ENG2	X					

No unexpected ENG1 IDLE (gear/slats)		X				
Respect gear/slats extension speeds			X			
Control ENG1 using gear/slats extensions		X				
Continue using ENG1 for approach/landing	X					
Identify A/THR as a problem/unreliable					X	
A/P ON after LS						X
Inform cabin crew of situation				X		
Event: Landing prep						
Check impact of failures on landing performance (LPC-NG / QRH)			X			
Discuss GA performance risk: ENG2 overheat, clean up for ENG1, Flap 3 landing			X	X		
request equipment on runway/engines off				X		

Appendix E

Chapter 3 DFCP and strategy phase behavioral measures data

Table E.1. DFCP scores

Group	Crew ID	Total DFCP	ENG2 Management	ENG1 Management	Route Management	Communication
BSL	102	67%	45%	71%	88%	83%
	103	57%	36%	57%	50%	67%
	202	70%	55%	71%	88%	67%
STG	101	73%	73%	57%	100%	67%
	104	70%	64%	86%	63%	67%
	106	77%	55%	86%	100%	33%
	107	63%	55%	86%	38%	67%
	108	67%	36%	71%	88%	67%
	109	43%	45%	14%	50%	67%
	110	60%	45%	71%	50%	33%
RIS	206	67%	36%	86%	88%	67%
	207	67%	45%	86%	88%	50%
	208	80%	64%	100%	88%	50%
	209	83%	73%	86%	100%	67%

Table E.2. Strategy phase duration (observed) scores

Group	Crew ID	P1 Duration (observed)	P2 Duration (observed)	P3 Duration (observed)	P4 Duration (observed)	P5 Duration (observed)	P6 Duration (observed)
BSL	102	00:03:07	00:03:00	00:15:43	00:33:09	00:40:20	00:33:55
	103	00:06:00	00:00:00	00:19:50	00:24:37	00:29:00	00:22:15
	202	00:08:59	00:03:20	00:15:13	00:24:47	00:33:00	00:13:25
STG	101	00:04:17	00:03:54	00:10:33	00:19:11	00:30:10	00:15:31
	104	00:23:47	00:06:00	00:23:12	00:15:01	00:24:20	00:29:08
	106	00:10:00	00:04:10	00:06:06	00:12:45	00:26:34	00:15:40
	107	00:12:06	00:06:07	00:26:23	00:16:01	00:13:37	00:09:04
	108	00:03:33	00:06:11	00:12:10	00:29:53	00:43:56	00:23:19
	109	00:10:54	00:06:25	00:19:44	00:14:59	00:26:03	00:09:00
	110	00:03:00	00:06:00	00:25:46	00:20:54	00:21:38	00:14:45
RIS	206	00:03:00	00:03:00	00:10:08	00:18:43	00:34:35	00:34:15
	207	00:13:49	00:03:00	00:13:56	00:18:09	00:49:45	00:31:08
	208	00:11:22	00:03:20	00:19:26	00:32:27	00:46:12	00:34:46
	209	00:13:35	00:03:21	00:19:46	00:32:45	00:46:28	00:36:01

Table E.3. Strategy phase duration (attention corrected) scores

Group	Crew ID	P1 Duration (attention corrected)	P2 Duration (attention corrected)	P3 Duration (attention corrected)	P4 Duration (attention corrected)	P5 Duration (attention corrected)	P6 Duration (attention corrected)
BSL	102	00:00:59	00:01:09	00:05:27	00:15:58	00:20:23	00:15:54
	103	00:02:05	00:00:00	00:08:39	00:14:53	00:16:46	00:15:23
	202	00:02:41	00:01:10	00:10:10	00:11:44	00:17:00	00:05:05
STG	101	00:01:23	00:01:13	00:03:50	00:09:01	00:17:45	00:09:17
	104	00:09:02	00:01:58	00:12:37	00:06:13	00:11:16	00:16:11
	106	00:04:09	00:01:27	00:02:22	00:08:44	00:15:45	00:11:07
	107	00:05:48	00:01:57	00:13:31	00:09:43	00:08:34	00:04:24
	108	00:00:47	00:01:49	00:06:59	00:18:19	00:28:05	00:14:48
	109	00:03:19	00:01:50	00:07:34	00:05:15	00:13:46	00:03:09
	110	00:00:50	00:02:01	00:11:47	00:07:54	00:08:25	00:06:45
RIS	206	00:01:09	00:00:48	00:03:22	00:06:28	00:16:31	00:16:09
	207	00:04:31	00:00:38	00:04:48	00:05:45	00:24:40	00:13:31
	208	00:03:32	00:01:04	00:09:12	00:13:30	00:20:18	00:15:06
	209	00:04:29	00:00:50	00:07:59	00:13:56	00:21:43	00:15:23

Table E.4. SC (total, normalized) behavioral measures scores

Group	Crew ID	SC1: Looping (total, normalized)	SC2: Sequence (total, normalized)	SC3: Skip 3 or 5 (total, normalized)	SC4: Back to 123 (total, normalized)	SC5: FFD 456 (total, normalized)
BSL	102	300%	123%	129%	21%	36%
	103	203%	125%	167%	34%	38%
	202	100%	97%	117%	26%	26%
STG	101	128%	73%	100%	26%	31%
	104	76%	74%	74%	41%	64%
	106	172%	65%	129%	9%	22%
	107	39%	118%	150%	63%	31%
	108	244%	250%	29%	23%	27%
	109	67%	49%	84%	23%	47%
	110	58%	82%	121%	27%	24%
RIS	206	72%	79%	23%	19%	36%
	207	94%	79%	69%	17%	39%
	208	160%	94%	79%	23%	45%
	209	175%	102%	117%	28%	41%

Table E.5. SC (total, count) behavioral measures scores

Group	Crew ID	SC1: Looping (count, normalized)	SC2: Sequence (count, normalized)	SC3: Skip 3 or 5 (count, normalized)	SC4: Back to 123 (count, normalized)	SC5: FFD 456 (count, normalized)
BSL	102	8.67	4.60	4.50	1.00	2.00
	103	5.00	2.80	5.50	1.42	1.38
	202	3.33	3.00	4.00	1.25	0.75
STG	101	3.67	2.40	3.00	1.17	1.25
	104	4.33	3.40	4.50	2.33	2.88
	106	4.00	1.60	4.00	0.83	0.63
	107	1.00	3.00	2.00	2.25	0.88
	108	5.67	5.60	1.00	1.58	1.13
	109	2.33	4.40	4.50	1.83	3.00
	110	4.00	4.40	4.50	1.67	2.13
RIS	206	3.67	3.60	2.00	1.08	2.00
	207	3.67	4.00	2.50	1.33	2.88
	208	8.00	4.40	2.50	1.33	2.00
	209	6.33	4.80	6.00	1.42	2.50

Appendix F

Chapter 3 Philosophy principles behavioral measures data

Table F.1. Principle duration (observed), principle duration (attention corrected) behavioral measures scores

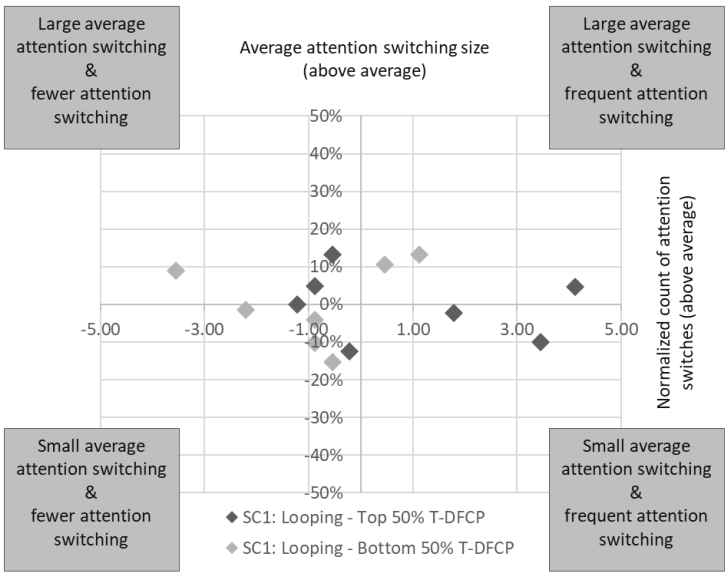
Group	Crew ID	TM Duration (observed)	UM Duration (observed)	CM Duration (observed)	TM Duration (attention corrected)	UM Duration (attention corrected)	CM Duration (attention corrected)
BSL	102	00:19:00	00:33:09	00:53:21	00:07:43	00:17:17	00:34:51
	103	00:20:15	00:24:37	00:46:40	00:10:05	00:15:10	00:32:32
	202	00:25:00	00:24:47	00:15:31	00:14:28	00:12:00	00:23:24
STG	101	00:12:29	00:19:11	00:39:47	00:05:37	00:09:44	00:27:07
	104	00:37:04	00:15:01	00:42:16	00:22:57	00:07:25	00:26:55
	106	00:11:45	00:12:45	00:37:25	00:07:10	00:08:58	00:29:03
	107	00:30:40	00:16:01	00:22:37	00:20:11	00:09:50	00:13:55
	108	00:13:28	00:29:53	00:55:28	00:08:30	00:19:00	00:43:16
	109	00:24:11	00:14:59	00:28:25	00:11:21	00:06:18	00:17:14
	110	00:26:27	00:20:54	00:29:10	00:13:37	00:08:51	00:15:14
RIS	206	00:14:16	00:18:43	00:00:00	00:05:35	00:07:49	00:15:56
	207	00:23:53	00:18:09	00:14:48	00:10:13	00:06:57	00:30:06
	208	00:29:17	00:32:27	00:14:48	00:15:10	00:15:06	00:23:41
	209	00:27:00	00:32:45	00:38:43	00:12:46	00:16:27	00:32:50

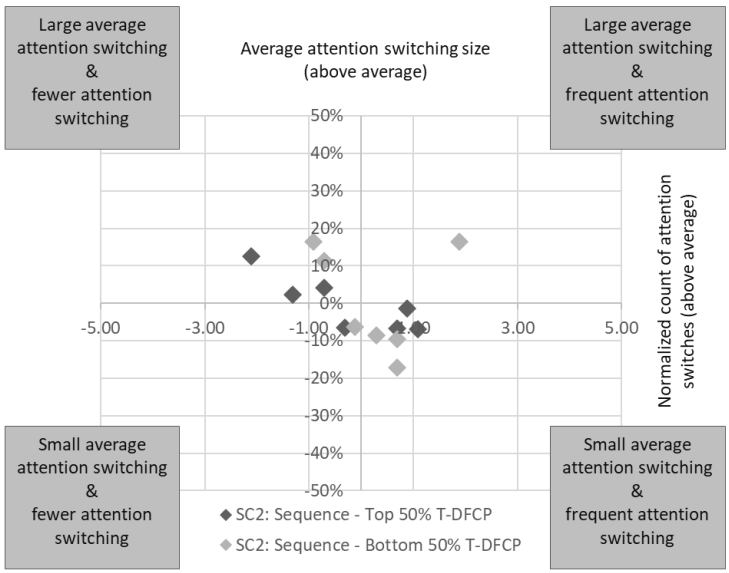
Table F.2. PC (total, normalized) and PC (count, normalized) behavioral measures scores

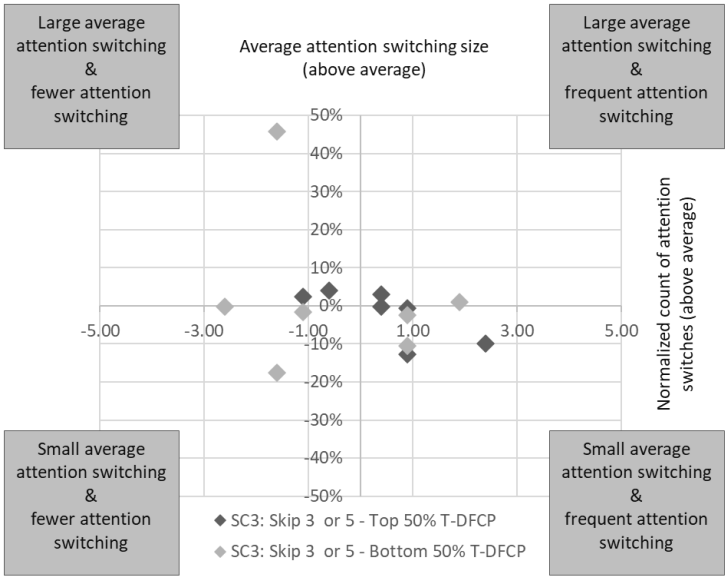
Group	Crew ID	PC1: Sequence and loop (total, normalized)	PC2: Return to TM (total, normalized)	PC3: Skip UM (total, normalized)	PC1: Sequence and loop (count, normalized)	PC2: Return to TM (count, normalized)	PC3: Skip UM (count, normalized)
BSL	102	196%	39%	50%	6.50	2.67	2.50
	103	183%	94%	92%	5.25	3.33	2.00
	202	123%	67%	113%	4.75	3.33	3.50
STG	101	98%	50%	96%	3.75	1.67	2.50
	104	71%	117%	133%	3.25	3.67	3.50
	106	100%	39%	42%	3.00	1.00	1.50
	107	104%	111%	67%	3.00	3.67	2.00
	108	283%	56%	25%	6.25	2.67	1.50
	109	42%	61%	117%	3.25	3.33	3.50
	110	102%	50%	38%	5.00	2.67	3.00
RIS	206	58%	28%	67%	2.00	2.33	2.00
	207	69%	44%	113%	3.75	3.33	4.50
	208	113%	78%	100%	5.25	2.67	4.00
	209	165%	72%	79%	6.75	2.00	3.00

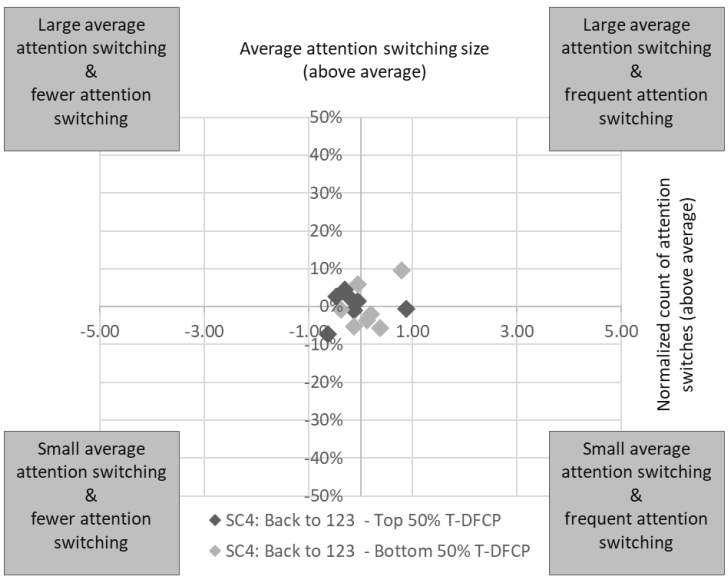
Appendix G

Characterization modalities analysis results






Large average attention switching & fewer attention switching
Average attention switching size (above average)
Large average attention switching & frequent attention switching
50%
40%
30%
20%
10%
0%
-10%
-20%
-30%
-40%
-50%
-5.00
-3.00
-1.00
1.00
3.00
5.00
Normalized count of attention switches (above average)
Small average attention switching & fewer attention switching
Small average attention switching & frequent attention switching
SC3: Skip 3 or 5 - Top 50% T-DFCP
SC3: Skip 3 or 5 - Bottom 50% T-DFCP


Large average attention switching & fewer attention switching
Average attention switching size (above average)
Large average attention switching & frequent attention switching
50%
40%
30%
20%
10%
0%
-10%
-20%
-30%
-40%
-50%
-5.00
-3.00
-1.00
1.00
3.00
5.00
Normalized count of attention switches (above average)
Small average attention switching & fewer attention switching
Small average attention switching & frequent attention switching
SC4: Back to 123 - Top 50% T-DFCP
SC4: Back to 123 - Bottom 50% T-DFCP

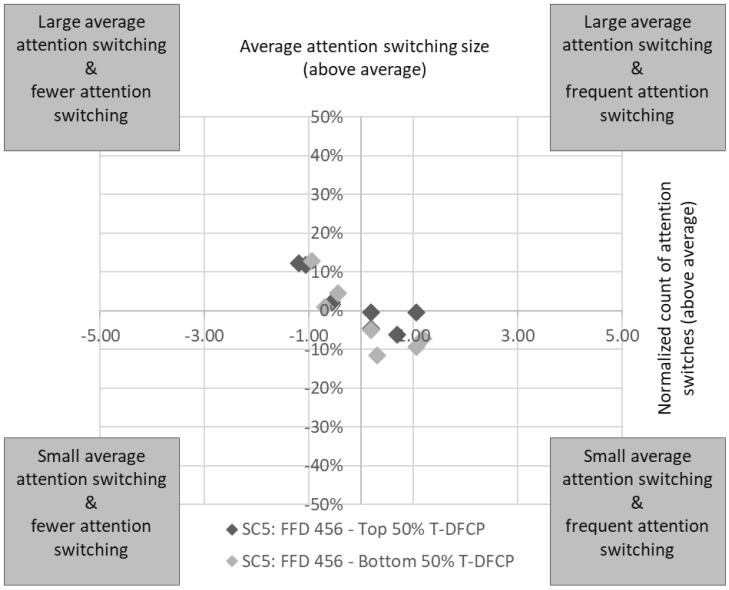
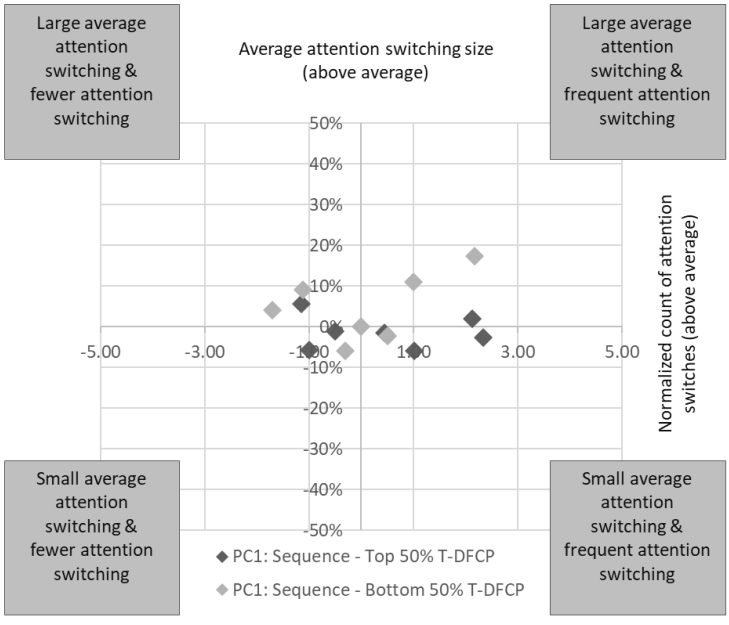
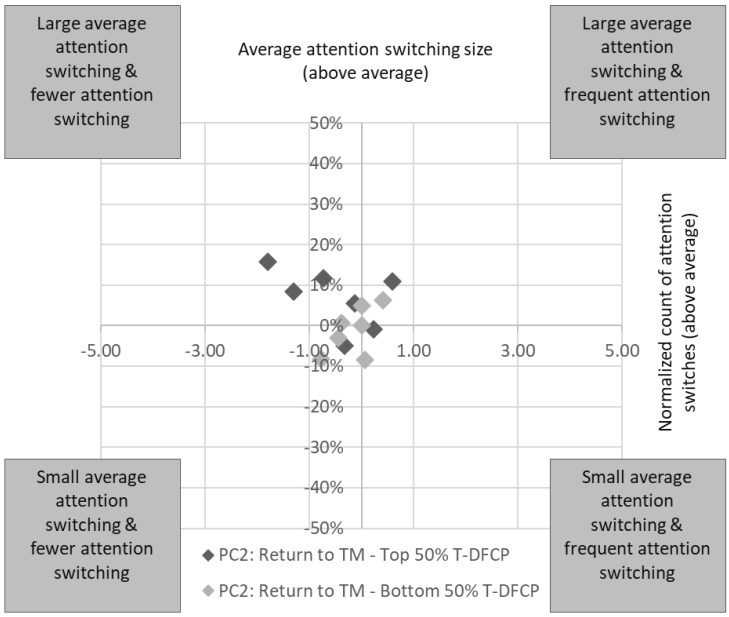


Figure G.1. Comparison of modalities for all five strategy phase characterizations indicating top and bottom half of performers (according to Total DFCP scores)





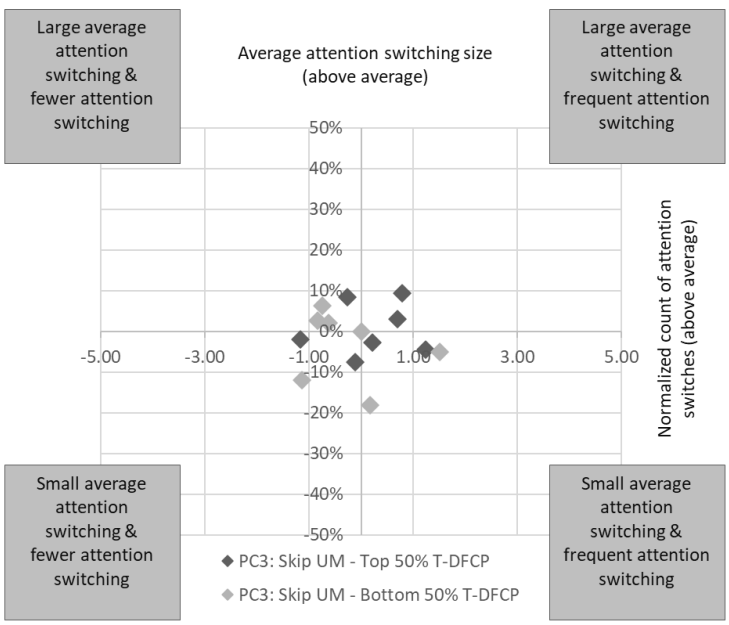


Figure G.2. Comparison of modalities for all three principle characterizations indicating top and bottom half of performers (according to Total DFCP scores)

Appendix H

Further considerations for training Airmanship 2.0

Pilot selection

Existing commercial pilot selection programs are generally quite skill-oriented, with a focus on basic psychological testing, handling skills and increasingly basic socialization skills being assessed. More advanced selection programs feature a competency-based approach, which tends to be a better predictor of pilot suitability (González Cabeza et al., 2021). Some air force selection programs have even reduced this to selecting solely on intelligence, personality and medical checks, as evidence shows that these are the most relevant predictors for trainability in military flight (Caretta et al., 2014). In a similar vein, how could selection programs enrich themselves further with useable indicators for problem-solving, mindfulness and emotional maturity? A possible improvement could be by cross-referencing metrics such as intelligence (a basic prerequisite), evaluation of behavioral markers in role playing exercises and interviews that explore a candidate's meta-cognitive awareness (e.g., self-awareness, tolerance for ambiguity, cognitive flexibility).

Ab-initio Training & MCC

While initial selection may only filter out those candidates which have a poor disposition from the onset, selected candidates are by no means competent in the cognitive abilities required. Furthermore, skills such as emotional regulation and mindfulness can take months or even years to manifest effectively in a person. Therefore, ab-initio training may be the best moment to begin developing these abilities. For example, mindfulness training adapted to the flight training context could be a valuable extension of the ground school program. Abilities such as non-judgmental awareness and startle recovery are so universal that they could even be developed prior to any flight training, with benefits for a candidate's general well-being to boot. Such training objectives could merge well with the shift to ab-initio CBTA programs such as the MPL program. For example, in the Multi-Crew Coordination (MCC) training program, case-based exercises could be used to reinforce CRM collaboration skills with shared awareness and conceptual sensemaking techniques. Also consider the value of developing emotional maturity to boost crew coordination training in the MCC phase.

Type Rating & ZFTT

The (initial) type rating is often the first time a candidate experiences the system and operational complexity requiring the new emotional and sensemaking abilities described in this chapter. This training phase provides an invaluable opportunity to connect these already trained and internalized abilities with the operational context that requires them. The technical and procedural nature of most type rating programs would have to make room for sensemaking cases which present the candidate with a situation that is *experienced* as ambiguous and opaque. Guided practice to apply these (meta-)cognitive abilities allows trainees to experience how these approaches structure their resolution of these cases. Better yet, consider that purposely presenting such cases before the candidate has internalized all the aircraft-specific knowledge and procedure training would more readily present an opaque situation (due to artificial lack of knowledge) and stimulate creative action.

More advanced problem-solving cases should also be provided near the end of the type rating or in the Zero Flight Time Training (ZFTT) phase (possibly as additional ground training) to allow the candidate to connect these problem-solving abilities with a complete knowledge of the aircraft and its procedures. Similar to other training phases in the regimen, CBTA type ratings (currently on the rise, with major OEM's developing them for their aircraft models) will mesh well with the non-technical nature of these (meta-) cognitive abilities.

Recurrent Training

Recurrent training has seen several marked developments and extensions in recent years which connect well with the (meta-)cognitive skill training, such as EBT. The CBTA competency frameworks can be readily expanded with additional competencies or observable behaviors (OB's) to encompass these (meta-)cognitive abilities as training objectives. Furthermore, the Line Oriented Flight Training (LOFT) based scenarios in such CBTA programs are also effective contexts to allow crews to experience a situation which develops in ambiguity and opacity. Refreshing practical startle and surprise recovery training (e.g., two-yearly) would also reinforce these skills without being very explicit about them, and CRM ground training can also feature exercises to practice collaborating in analytical and intuitive problem solving. A more radical proposal could be to extend the annual medical check with a mental-wellbeing check. This requires a significant shift towards establishing a safe environment for pilots to be honest about their mental health. A possible approach would be to shift from a mental "check" to mental coaching/training, to determine what type of support a pilot needs to improve his/her mental health (e.g., improve factors such as health, family, stressors, financial pressure).

Command Upgrade Training

Command training already focuses on specific competencies such as leadership, teamwork and decision making. However, given the differentiation between novice and

expert problem-solving strategies (analytical and intuitive, respectively), command training could be augmented to extend a candidate's problem-solving skillset with intuitive thinking and attention switching (increasing the ACC's activity). Such training could consist of advanced mindfulness training that address the reduction of self-doubt, which maintaining attention switching abilities and increased sensitivity to one's intuition. Furthermore, the command training should also ensure that crew captains are able to stimulate and facilitate an atmosphere conducive to healthy cognitive functioning and effective crew problem solving dynamics.

Instructor Training

Instructor qualifications and training have also improved significantly with the introduction of CBTA and EBT training programs. In particular, the Instructor & Examiner Competency (IEC) framework shared by ICAO, EASA and IATA (ICAO, 2023) is a good example where special emphasis has been placed on interpersonal skills, ability to facilitate and to manage a healthy learning environment. As such, instructors can be particularly useful to reinforce the Airmanship 2.0 (meta-)cognitive by facilitating self-reflection, and challenging crews at the appropriate level by understanding how opacity and ambiguity can be tuned to challenge the crews learning abilities. At a minimum, instructors should be competent in both analytical and intuitive problem-solving approaches, and may even have to put them to use to be able to develop a trainee's abilities with specific meditation exercises, *Socratic questioning* and personal interviews. This may restrict relatively young first officers to function as (type-rating) instructors.

Examiner Training

The examiner is a key agent to assure the quality of flight crews throughout all training phases. Where CBTA assessments are a challenge for today's instructors, assessing a candidate's mindfulness ability or intuition will be even more difficult as their abilities will not always feature clear external behavioral markers. Similar to instructors, examiners must be expert facilitators such that both the examiner and the candidate can discover the candidate's cognitive activities and assess if they are up to standard. In most cases, an examiner will have gained sufficient experience as an instructor to develop the necessary facilitation skills.

Many suggested training interventions described in this section are unconventional when compared to the task- and procedure-centered training that constitutes the mainstay of today's aviation training and checking. There is no doubt that these changes will create uncomfortable, confrontational experiences and encounter resistance and skepticism. However, experience in training and introducing CBTA and EBT within the industry has shown that – when framed appropriately – such changes are seen as a welcome, refreshing

change to the monotony and compliance it (partially) replaces. While EBT and CBTA are still relatively close to the operational context, the introduction of startle training, metacognition, mindfulness and higher order neurocognitive and self-reflective exercises will require trainees and trainers to explore training that is less operationally specific and much more personal in nature.

Appendix I

Mindfulness Surveys

Initial & Exit survey (before training, and after simulator)

Demographic header (birthyear, gender, role/rank, experience, recency)

Please indicate to what extent you agree or disagree with each of the following statements:

Strongly disagree / Disagree / Somewhat disagree / Somewhat agree / Agree / Strongly agree

I have a certain amount of skills required in my job (such as technical skills or CRM skills) and I really can't do much to change it.

Skills required in my job are something about me that I can't change much.

I can learn new things, but I can't really change my basic skills relevant to my job.

I am confident I get the success I deserve in life.

Sometimes I feel depressed.

When I try, I generally succeed.

Sometimes when I fail I feel worthless.

I complete tasks successfully.

Sometimes, I do not feel in control of my work.

Overall, I am satisfied with myself.

I am filled with doubts about my competence.

I determine what will happen in my life.

I do not feel in control of my success in my career.

I am capable of coping with most of my problems.

There are times when things look pretty bleak and hopeless to me.

Please read the following statements and indicate to what extent each statement describes yourself:

Not at all like me / Not like me / Somewhat like me / Like me / Very much like me

I often set a goal but later choose to pursue a different one.

New ideas and new projects sometimes distract me from previous ones.

I become interested in new pursuits every few months.

My interests change from year to year.

I have been obsessed with a certain idea or project for a short time but later lost interest.

I have difficulty maintaining my focus on projects that take more than a few months to complete.

I have achieved a goal that took years of work.

I have overcome setbacks to conquer an important challenge.

I finish whatever I begin.

Setbacks don't discourage me.

I am a hard worker.

I am diligent.

Please take a moment to think about what makes your life feel important to you. Please respond to the following statements as truthfully and accurately as you can, and also please remember that these are very subjective questions and that there are no right or wrong answers. Please answer according to the scale below:

Absolutely untrue / Mostly untrue / Somewhat untrue / Can't say true or false / Somewhat true / Mostly true / Absolutely true

I understand my life's meaning.

I am looking for something that makes my life feel meaningful.

I am always looking to find my life's purpose.

My life has a clear sense of purpose.

I have a good sense of what makes my life meaningful.

I have discovered a satisfying life purpose.

I am always searching for something that makes my life feel significant.

I am seeking a purpose or mission for my life.

My life has no clear purpose.

I am searching for meaning in my life.

Please indicate how much support in performing your work you currently receive from each of three types of people.

Absolutely unsupportive / Mostly supportive / Somewhat unsupportive / Somewhat supportive / Mostly supportive / Absolutely supportive

Peers

Direct supervisor (e.g., training/standards manager, fleet captain, etc.)

Superiors (which, in this organization, refers to upper management)

Please indicate to what extent you agree or disagree with each of the following statements:

Strongly disagree / Disagree / Somewhat disagree / Somewhat agree / Agree / Strongly agree

I take risks that bring excitement into my life.

I often think about how things were earlier in my life.

I am usually certain about what I am going to do next.

I put off small gratifications 1 can get now in order to try for bigger gratifications later.

I like to be spontaneous and make decisions on the spur of the moment.

I try to be realistic about what the future holds for me.

I try to live one day at a time.

Sometimes I wish I could go back to relive or change my past experience(s).

It's more important for me to enjoy what I am doing than it is to get things done "on time"

I prefer the old, familiar, and known ways of doing things to new and changing ways.

Planning activities takes all the fun out of them.

My plans about the future are pretty well laid out.

My behavior seems to be more influenced by past experiences than by future goals.

Most of my thoughts are about things that have already happened.

I don't think much about what did or will happen, only about what is happening now.

What I do today is focused on making tomorrow better.

I live to experience what is, rather than worrying about what will be.

When someone hurts or angers me, it is hard for me to forgive and forget.

The best way to do things well is to take them as they come.

When I want to accomplish something, I set goals and consider means for achieving them.

How I behave today is a direct reflection of my past experiences.

I am able to resist temptations when I know there is work to be done.

I believe it is important to save for a rainy day.

I often feel as though I were reliving experiences from my past.

If I can't see the immediate benefits of doing something, I won't do it.

I think about the future consequences of my actions.

I often talk about my past experiences with others.

It is best to live day-to-day and let tomorrow take care of itself.

Please indicate to what extent you agree or disagree with each of the following statements:

Strongly disagree / Disagree / Somewhat disagree / Somewhat agree / Agree / Strongly agree

I tend to bounce back quickly after hard times.

I have a hard time making it through stressful events.

It does not take me long to recover from a stressful event.

It is hard for me to snap when something bad happens.

I usually come through difficult times with little trouble.

I tend to take a long time to get over setbacks in my life.

Please indicate to what extent you agree or disagree with each of the following statements:

Strongly disagree / Disagree / Somewhat disagree / Somewhat agree / Agree / Strongly agree

I find real enjoyment in my job.

I like my job better than the average person.

I am seldom bored with my job.

I would not consider taking another kind of job.

Most days I am enthusiastic about my job.

I Feel Fairly well satisfied with my job.

Below is a collection of statements about your everyday experience. Please indicate how infrequently or frequently you currently have each experience. Please answer according to what really reflects your experience rather than what you think your experience should be. Please treat each item separately from every other item.

Almost never / Very infrequently / Somewhat infrequently / Somewhat frequently / Very frequently / Almost always

I could be experiencing some emotion and not be conscious of it until some time later.

I break or spill things because of carelessness, not paying attention, or thinking of something else.

I find it difficult to stay focused on what's happening in the present

I tend to walk quickly to get where I'm going without paying attention to what I experience along the way.

I tend not to notice feelings of physical tension or discomfort until they really grab my attention.

I forget a person's name almost as soon as I've been told it for the first time.

It seems I am "running on automatic," without much awareness of what I'm doing.

I rush through activities without being really attentive to them.

I get so focused on the goal I want to achieve that I lose touch with what I'm doing right now to get there.

I do jobs or tasks automatically, without being aware of what I'm doing.

I find myself listening to someone with one ear, doing something else at the same time.

I drive places on "automatic pilot" and then wonder why I went there.

I find myself preoccupied with the future or the past.

I find myself doing things without paying attention.

I snack without being aware that I'm eating.

Operational Survey (after every flight, voluntary)

Please indicate to what extent do you agree with each of the following statements with respect to your flight operation experience:

Strongly disagree / Disagree / Somewhat disagree / Somewhat agree / Agree / Strongly agree

I am able to block out most other distractions.

I am absorbed in what I am doing.

I am immersed in the task I am performing.

I get distracted by other attentions very easily.

My attention does not get diverted very easily.

Please answer the following questions:

None / A little / Some amount / A lot / Extreme amount

How much self-control do you feel right now?

How much willpower do you feel right now?

Please indicate to what extent do you believe the communication in the cockpit was _____ during the flight operation

Choose a number between 0 and 10 (not at all) and 10 (extremely):

Adequate: _____

Timely: _____

Accurate: _____

Complete: _____

Credible: _____

Please indicate to what extent each of the following describes yourself in the cockpit.

Not at all like me / Not like me / Somewhat like me / Like me / Very much like me

Identifies others' strengths and limitations.

Accurately reads people's moods, feelings, or nonverbal cues.

Gives others opportunity to speak their mind.

Accurately assesses the underlying or root causes of a person's problems.

Pays attention and listens well.

Shows sensitivity and understanding.

Asks questions to be sure he/she understands another person.

Please answer the following question:

None / Rarely / Sometimes / Most of the time / All the time

How much of the time during the past week did you feel happy?

How things were going in your life during the past week? Choose a number between 0 and 10 (0: the worst possible life you could imagine to 10: the best possible life you could imagine) _____

How much of the time during the past week did you _____?

None / Rarely / Sometimes / Most of the time / All the time

feel worn out

feel tired

feel downhearted and blue

feel calm and peaceful

feel nervous

have a lot of energy

feel so down in the dumps that nothing could cheer you up

Please indicate to what extent your personal life or relationship affected the following items during the past week.

Not at all / To a small extent / To some extent / To a large extent / To a very large extent

Work performances

Morale

Teamwork

References

Abbott, K. H. (2017). Human factors engineering and flight deck design. *Digital avionics handbook*, 302-328.

ACROSS consortium. (2016). Final Report Summary – ACROSS (Advanced Flight deck for Reduction Of StreSs and workload).

Adams-White, J., Wheatcroft, J. M., & Jump, M. (2017). *Measuring the Meta and Cognitive abilities of Air Defence Operators* (Vol. 20).

Adanov, L., Efthymiou, M., & Macintyre, A. (2020). An exploratory study of pilot training and recruitment in Europe. *International Journal of Aviation Science and Technology*, *1*(02), 44-51.

Air Data News. (2025). French prosecutor pushes for Airbus and Air France convictions in A330 AF447 disaster. *Air Data News*. https://www.airdatanews.com/french-prosecutor-pushes-for-airbus-and-air-france-convictions-in-a330-af447-disaster/

Airbus. (2025). A Statistical Analysis of Commercial Aviation Accidents 1958 - 2024. Blagnac, France.

Ajzen, I. (1991). The theory of planned behavior. Organizational Behavior and Human Decision Processes, 50 (2): 179-211.

Akinci, C., & Sadler-Smith, E. (2020). 'If something doesn't look right, go find out why': how intuitive decision making is accomplished in police first-response. *European Journal of Work and Organizational Psychology*, 29(1), 78–92.

Alaydi, B., Ng, S. I., & Lim, X. J. (2025). The Role of Air Traffic Controllers' Mindfulness in Enhancing Air Traffic Safety: JDR Theory in the Saudi Arabian Aviation Context. *Logistics*, 9(3).

Aldous, C. R. (2007). Creativity, problem solving and innovative science: Insights from history, cognitive psychology and neuroscience. *International Education Journal*, 8(2), 176–186.

Andersson, L., Eriksson, J., Stillesjö, S., Juslin, P., Nyberg, L., & Wirebring, L. K. (2020). Neurocognitive processes underlying heuristic and normative probability judgments. *Cognition*, 196.

Austin, J. H. (2013). Zen and the brain: mutually illuminating topics. *Frontiers in Psychology*, 4.

Australian Transport Safety Bureau (ATSB). (2011). In-flight upset - 154 km west of Learmonth, WA, 7 October 2008, VH-QPA, Airbus A330-303 (Investigation No. AO-2008-070). atsb.gov.au

Australian Transport Safety Bureau (ATSB). (2013). In-flight uncontained engine failure: Airbus A380-842, VH-OQA, overhead Batam Island, Indonesia, 4 November 2010 (Investigation No. AO-2010-089). atsb.gov.au

Australian Transportation Safety Bureau (ATSB). (2013). In-flight uncontained engine failure Airbus A380-842, VH-OQA.

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

Banks, V. A., Plant, K. L., & Stanton, N. A. (2020). Leaps and shunts: designing pilot decision aids on the flight deck using Rasmussen's ladder. *Contemporary ergonomics and human factors.*

Barrouillet, P. (2011). Dual-process theories and cognitive development: Advances and challenges. *Developmental Review* (Vol. 31, Issues 2–3, pp. 79–85).

Beaty, R. E., Benedek, M., Wilkins, R. W., Jauk, E., Fink, A., Silvia, P. J., Hodges, D. A., Koschutnig, K., & Neubauer, A. C. (2014). Creativity and the default network: A functional connectivity analysis of the creative brain at rest. *Neuropsychologia*, 64, 92–98.

Berkovich-Ohana, A., Glicksohn, J., Ben-Soussan, T. D., & Goldstein, A. (2017). Creativity Is Enhanced by Long-Term Mindfulness Training and Is Negatively Correlated with Trait Default-Mode-Related Low-Gamma Inter-Hemispheric Connectivity. *Mindfulness*, 8(3), 717–727.

Billings, C. E. (1996). *Aviation automation: The search for a human-centered approach*. Lawrence Erlbaum Associates.

Bjurling, O., Müller, H., Burgén, J., Bouvet, C. J., & Berberian, B. (2024). Enabling human-autonomy teaming in aviation: A framework to address human factors in digital assistants design. *Journal of Physics: Conference Series* (Vol. 2716, No. 1, p. 012076). IOP Publishing.

Blalock, S. J., & Reyna, V. F. (2016). Using Fuzzy-Trace Theory to Understand and Improve Health Judgments, Decisions, and Behaviors: A Literature Review. *Health Psychology*, 35(8), 781–792.

Boeing, (2019). Pilot & Technician Outlook 2019-2038.

Boeing. (2019). Boeing statement on AOA disagree alert [Press release]. https://boeing.mediaroom.com/news-releases-statements?item=130431

Bowles, G. (2018). Automation Readiness Levels. Personal Communication.

Boy, G. A. (2020). Aerospace human systems integration: Evolution over the last 40 years. *A framework of human systems engineering: Applications and case studies*, 113-128.

Bradshaw, J. M., Hoffman, R. R., Woods, D. D., & Johnson, M. (2013). The seven deadly myths of" autonomous systems". *IEEE Intelligent Systems*, *28*(3), 54-61.

Brand, M., Grabenhorst, F., Starcke, K., Vandekerckhove, M. M. P., & Markowitsch, H. J. (2007). Role of the amygdala in decisions under ambiguity and decisions under risk: Evidence from patients with Urbach-Wiethe disease. *Neuropsychologia*, 45(6), 1305–1317.

Bremer, B., Wu, Q., Mora Álvarez, M. G., Hölzel, B. K., Wilhelm, M., Hell, E., Tavacioglu, E. E., Torske, A., & Koch, K. (2022). Mindfulness meditation increases default mode, salience, and central executive network connectivity. *Scientific Reports*, 12(1).

Brewer, J. A., Worhunsky, P. D., Gray, J. R., Tang, Y. Y., Weber, J., & Kober, H. (2011). Meditation experience is associated with differences in default mode network activity and connectivity. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50), 20254–20259.

Brezina, I. (2011). Unfolding intuition: Zen tradition in cognitive psychology lenses. *Studia Psychologica*, 53(2), 215.

Buch, J. P., Niedermeier, D., & Stepniczka, I. (2017). Managing the Unexpected-Human-in-the-Loop Simulation as Effective Tool for the Assessment of the Risk Information System in an Operationally Relevant Context. *AIAA Modelling and Simulation Technologies Conference* (p. 4155).

Bureau d'Enquêtes et d'Analyses (BEA). (2012). Final Report On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZC operated by Air France flight AF 447 Rio de Janeiro – Paris.

Byrne, E. K., & Thatchenkery, T. (2019). Cultivating creative workplaces through mindfulness. *Journal of Organizational Change Management*, 32(1), 15–31.

Cahill, J., Callari, T. C., Fortmann, F., Suck, S., Javaux, D., Hasselberg, A. & van Doorn, B. A. (2017). Adaptive automation and the third pilot: managing teamwork and workload in an airline flight deck. *International Symposium on Human Mental Workload: Models and Applications* (pp. 161-173). Cham: Springer International Publishing.

Calabretta, G., Gemser, G., & Wijnberg, N. M. (2017). The Interplay between Intuition and Rationality in Strategic Decision Making: A Paradox Perspective. *Organization Studies*, 38(3–4), 365–401.

Caldwell, J. A. (2012). Crew schedules, sleep deprivation, and aviation performance. *Current Directions in Psychological Science*, 21(2), 85-89.

Calhoun, G. (2022). Adaptable (not adaptive) automation: Forefront of human–automation teaming. *Human factors*, 64(2), 269-277.

Caraway, C. L. (2020). A looming pilot shortage: It is time to revisit regulations. *International Journal of Aviation, Aeronautics, and Aerospace, 7(2), 3*.

Carleton, T., Cockayne, W., & Leifer, L. (2007). An Exploratory Study about the Role of Ambiguity During Complex Problem Solving.

Carretta, T. R., Teachout, M. S., Ree, M. J., Barto, E. L., King, R. E., & Michaels, C. F. (2014). Consistency of the relations of cognitive ability and personality traits to pilot training performance. *The International Journal of Aviation Psychology*, 24(4), 247–264.

Carruthers, P. (2007). The creative-action theory of creativity. *The innate mind: Foundations and the future*. Oxford University Press.

Casner, S. M., Geiselman, E. E., & Morrow, D. G. (2014). The effectiveness of airline pilot training for abnormal events. *Human Factors*, 56(8), 1390-1402.

Causse, M., Mercier, M., Lefrançois, O., & Matton, N. (2025). Impact of automation level on airline pilots' flying performance and visual scanning strategies: A full flight simulator study. *Applied Ergonomics*, 125, 104456.

Chant, C. (2002). *A Century of Triumph: The History of Aviation*. Simon and Schuster.

Charpentreau, C. (2025). Boeing reaches controversial settlement, avoids criminal trial over MAX crashes. *AeroTime*. https://www.aerotime.aero/articles/boeing-reaches-controversial-settlement-avoids-criminal-trial-over-max-crashes

Chassy, P., & Gobet, F. (2011). A Hypothesis About the Biological Basis of Expert Intuition. *Review of General Psychology*, 15(3), 198–212.

Chen, J. (2025). Quantifying Startle and Surprise: Development of Measuring Instruments and Validation in an Aviation Context.

Chokshi, N. (2025). U.S. government admits liability in deadly D.C.-area air collision. T*he New York Times*. https://www.nytimes.com/2025/12/17/us/politics/dc-plane-crash-government-liability.html

Clark, A. N., & Wilson, A. L. (2024). From Levers to Glass: The Evolution of Aircraft Flight decks and the Future of Aviation Innovation.

Cohen, M. X., Heller, A. S., & Ranganath, C. (2005). Functional Connectivity with Anterior Cingulate and Orbitofrontal Cortices During Decision-Making.

Colautti, L., Antonietti, A., & Iannello, P. (2022). Executive Functions in Decision Making under Ambiguity and Risk in Healthy Adults: A Scoping Review Adopting the Hot and Cold Executive Functions Perspective. *Brain Sciences* (Vol. 12, Issue 10). MDPI.

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: Frontal lobe function and human decision-making. *PLoS Biology*, 10(3).

Constant II, E. W. (1973). A model for technological change applied to the turbojet revolution. *Technology and Culture, 14(4),* 553-572.

Conway-Smith, B., & West, R. L. (2022). System-1 and System-2 realized within the Common Model of Cognition.

Crouch, V. (2020). Analysis of the airline pilot shortage. *Scientia et Humanitas,* 10, 93-106.

Cullen, P., Cahill, J., & Gaynor, K. (2021). A qualitative study exploring well-being and the potential impact of work-related stress among commercial airline pilots. *Aviation Psychology and Applied Human Factors.*

Curran, R., & Pulver, R. (2024). The Impact of Reduced Flying Experience on Airline Pilots' Proficiency During the COVID-19 Pandemic. *AIAA Aviation Forum and Ascend 2024* (p. 3947).

Dane, E., & Pratt, M. G. (2007). Exploring intuition and its role in managerial decision making.

de Brito, G., & Boy, G. (1999). Situation awareness and procedure following. *CSAPC* (Vol. 99, pp. 9-14).

De Crespigny, R. (2012). *QF32*. Macmillan.

De Houwer, J. (2019). Moving Beyond System 1 and System 2. *European Psychologist*, 14(3), 260–262.

De Neys, W., & Pennycook, G. (2019). Logic, Fast and Slow: Advances in Dual-Process Theorizing.

Degani, A., & Wiener, E. L. (1990). Cockpit checklists: Concepts, design, and use (NASA Contractor Report No. 177549). NASA Ames Research Center. nasa.gov

Degani, A., & Wiener, E. L. (1994). *On the design of flight-deck procedures* (No. A-94095).

Dekker, S. W. (2002). Reconstructing human contributions to accidents: the new view on error and performance. *Journal of safety research*, 33(3), 371-385.

Dekker, S. W., & Lundstrom, J. (2006). From threat and error management (TEM) to resilience. *Human Factors and Aerospace Safety*, 6(3), 261.

Dennin, A., Furman, K., Pretz, J. E., & Roy, M. M. (2022). The relationship of types of intuition to thinking styles, beliefs, and cognitions. *Journal of Behavioral Decision Making*, 35(5).

DeSalvo, P., Fogarty, D. (2016). DOT/FAA/TC-16/39 Safety Issues and Shortcomings with Requirements Definition, Validation and Verification Processes – Final Report.

Diehl, A. E. (2013). Air safety investigators: using science to save lives—one crash at a time. Xlibris Corporation.

Dijkstra, A. (2023). Safety as Airline Business Aspect: From Data to Action by A Value Model for Big Data and Feedback Method for Small FlightStories.

Dörfler, V., & Ackermann, F. (2012). Understanding intuition: The case for two forms of intuition. *Management Learning*, 43(5), 545–564.

Dreyfus, S. E. (2014). System 0: The overlooked explanation of expert intuition. *Handbook of research methods on intuition*. Edward Elgar Publishing.

Duchevet, A., Vo, D. B., Peyruqueou, V., De-La-Hogue, T., Garcia, J., Causse, M., & Imbert, J. P. (2024). FOCUS: An intelligent startle management assistant for maximizing pilot resilience. *ICCAS24* (pp. 5-12). SCITEPRESS-Science and Technology Publications.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(1), 32-64.

Endsley, M., & Jones, W. M. (1997). Situation Awareness Information Dominance & Information Warfare. Logicon Technical Services Inc. Dayton, OH.

Enea, G., & Porretta, M. (2012, September). A comparison of 4D-trajectory operations envisioned for Nextgen and SESAR, some preliminary findings. 28th Congress of the International Council of the aeronautical sciences (Vol. 5, pp. 4152-4165). Edinburgh, UK: Optimage Ltd.

Epstein, S. (2010). Intuition and decision-making. *Intuição e decisão*, 29.

Ethiopia Aircraft Accident Investigation Bureau (EAIB). (2022). Investigation report on accident to the B737-MAX8 reg. ET-AVJ operated by Ethiopian Airlines.

Eurocontrol. (2020). The FLY AI Report: Demystifying and accelerating AI in Aviation/ATM.

European Aviation Safety Agency (EASA). (2013). Annual safety review 2013.

European Aviation Safety Agency (EASA). (2019). Study on "atypical" employment in aviation.

European Union Aviation Safety Agency (EASA). (2020). Easy access rules for flight crew licencing (Part-FCL) (June 2020 version).

European Union Aviation Safety Agency (EASA). (2023). Artificial intelligence roadmap 2.0.: Human-centric approach to AI in aviation. Version 2.0.

European Union Aviation Safety Agency (EASA). (2023). Certification Standards Part 25.

Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241.

Fabio, R. A., & Towey, G. E. (2018). Long-term meditation: the relationship between cognitive processes, thinking styles and mindfulness. *Cognitive Processing*, 19(1), 73–85. h

Fadde, P. J., & Klein, G. A. (2010). Deliberate performance: Accelerating expertise in natural settings. *Performance Improvement*, 49(9), 5–14.

Farb, N. A. S., Segal, Z. v., Mayberg, H., Bean, J., Mckeon, D., Fatima, Z., & Anderson, A. K. (2007). Attending to the present: Mindfulness meditation reveals distinct neural modes of self-reference. *Social Cognitive and Affective Neuroscience*, 2(4), 313–322.

Feary, M. (2018). A first look at the evolution of flight crew requirements for emerging market aircraft (No. ARC-E-DAA-TN57004).

Federal Aviation Administration. (2024, July 25). Roadmap for artificial intelligence safety assurance (Version I). U.S. Department of Transportation.

Field, J., Fucke, L., Correia Grácio, B., & Mohrmann, J.F.W. (2016). Flight crew response to unexpected events: a simulator experiment. *AIAA Modelling and Simulation Technologies Conference* (p. 3373).

Field, J., Rankin, A., Mohrmann, J.F.W., Boland, E., & Woltjer, R. (2017). Flexible procedures to deal with complex unexpected events in the cockpit. *Resilience Engineering Association Symposium Liége, Belgum, 26-29 June, 2017*. Resilience Engineering Association.

Field, J.N., Boland, E.J., Van Rooij, J.M., Mohrmann, J.F.W., Smeltink, J.W. (2018). Startle Effect Management. (Report Nr. NLR-CR-2018-242). European Union Aviation Safety Agency.

Fishbein, M., & Ajzen, I. (2011). *Predicting and changing behavior: The reasoned action approach*. Taylor & Francis.

Flight Safety Foundation (FSF). (2026). 2023 safety report: Commercial airliner interactive dashboard. Aviation Safety Network. https://aviation-safety.net/dashboard/safetyreport2023

Flin, R., Martin, L., Goeters, K. M., Hormann, H. J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety*, 3, 97-120.

Fraher, A. L. (2019). The vulnerability of quasi-professional experts: A study of the changing character of US airline pilots' work. *Economic and Industrial Democracy*, 40(4), 867-889.

FutureSky Safety P6 Consortium (2016). D6.4 Recommendations recovery measures and HMI implementation.

Gago, C. P., Hansman, R. J., Edmondson, M. K., & Mosqueda, M. A. (2025). A Study of Pilot Response to System Failure in Transport Category Aircraft from 2000 to 2024. *2025 AIAA DATC/IEEE 44th Digital Avionics Systems Conference (DASC)* (pp. 1-8). IEEE.

Garcia, M. (2025). Air France flight 447: 16 years on, appeal reopens questions of accountability. *Aerospace Global News*. https://aerospaceglobalnews.com/news/af447-appeal-air-france-airbus/

Gardini, K. C., Chiavelli, F. H., Scarpari, J. R. S., & de Andrade, D. (2022). Conceptual models of startle, surprise and automation bias analyzed through recent aviation accident reports. *Journal of Aerospace Technology and Management*, *14*, e3022.

Garrison, K. A., Zeffiro, T. A., Scheinost, D., Constable, R. T., & Brewer, J. A. (2015). Meditation leads to reduced default mode network activity beyond an active task. *Cognitive, Affective and Behavioral Neuroscience*, 15(3), 712–720.

George, J. M., & Dane, E. (2016). Affect, emotion, and decision making. *Organizational Behavior and Human Decision Processes*, 136, 47–55.

Glöckner, A., & Witteman, C. (2010). Beyond dual-process models: A categorisation of processes underlying intuitive judgement and decision making. *Thinking and Reasoning*, 16(1), 1–25.

González Cabeza, I., Molesworth, B., Good, M., Caponecchia, C., & Steffensen, R. (2021). Investigating the predictive validity of the COMPASS pilot selection test. *The International Journal of Aerospace Psychology*, 31(3), 252–268.

Gorter, D. and Jaeger, C. (2014). Project Samurai Pilot. Cognitive Enhancement of Airline Pilots with Biofeedback Training. Delft University of Technology, Faculty of Aerospace Engineering, and University of Leiden, Faculty of Psychology.

Hancock, P. A., Cruit, J., Kochan, J. A., Kaplan, A. D., Diaz, Y., & Pruchnicki, S. (2022). Pilots' responses to unexpected events: Conceptual, theoretical, methodological, and analytical issues. *The International Journal of Aerospace Psychology*, *32*(4), 254-282.

Hannah, T. C., Turner, D., Kellner, R., Bederson, J., Putrino, D., & Kellner, C. P. (2022). Neuromonitoring Correlates of Expertise Level in Surgical Performers: A Systematic Review. *Frontiers in Human Neuroscience* (Vol. 16). Frontiers Media S.A.

Harris, D. (2017). *Human factors for civil flight deck design*. Routledge.

Harris, D. (2023). Single-pilot airline operations: Designing the aircraft may be the easy part. *The aeronautical journal*, *127*(1313), 1171-1191.

Hart, S. G. (2006). NASA-task load index (NASA-TLX); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 9, pp. 904-908). Sage Publications.

Haslbeck, A., Kirchner, P., Schubert, E., & Bengler, K. (2014). A flight simulator study to evaluate manual flying skills of airline pilots. *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 58, No. 1, pp. 11-15). Sage CA: Los Angeles, CA: SAGE Publications.

Helmreich, R. L., Klinect, J. R., & Wilhelm, J. A. (1999). Models of threat, error, and CRM in flight operations. *Proceedings of the tenth international symposium on aviation psychology* (Vol. 10, pp. 677-682).

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (2017). The evolution of crew resource management training in commercial aviation. *Human Error in Aviation* (pp. 275-288). Routledge.

Henley, I. M. (2017). *Aviation education and training: Adult learning principles and teaching strategies*. Routledge.

Hill, O. W. (1987). Intuition: Inferential Heuristic or Epistemic Mode? *Imagination, Cognition and Personality*, 7(2), 137–154.

Ho, N., Johnson, W., Panesar, K., Wakeland, K., Sadler, G., Wilson, N., & Brandt, S. (2017). Application of human-autonomy teaming to an advanced ground station for reduced crew operations. 2017 IEEE/AIAA 36th digital avionics systems conference (DASC) (pp. 1-4). IEEE.

Hodgkinson, G. P., Langan-Fox, J., & Sadler-Smith, E. (2008). Intuition: A fundamental bridging construct in the behavioural sciences. *British Journal of Psychology*, 99(1), 1–27.

Hogarth, R. M. (2014). Deciding analytically or trusting your intuition? The advantages and disadvantages of analytic and intuitive thought. *The routines of decision making* (pp. 97–112). Psychology Press.

Hollnagel, E. (2002). Understanding accidents-from root causes to performance variability. *Proceedings of the 2002 IEEE 7th conference on Human factors and power plants* (pp. 1-1). IEEE.

Hollnagel, E. (2014). *Safety-I and Safety-II: The past and future of safety management*. Ashgate.

Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. CRC Press.

Hölzel, B. K., Lazar, S. W., Gard, T., Schuman-Olivier, Z., Vago, D. R., & Ott, U. (2011). How does mindfulness meditation work? Proposing mechanisms of action from a conceptual and neural perspective. *Perspectives on Psychological Science* (Vol. 6, Issue 6, pp. 537–559).

Hörhager, P. (2025). Aircraft Automation and Human Performance: An Analysis of Flight Deck Failures. Doctoral dissertation, Technische Universität Wien.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). John Wiley & Son.

Hourlier, S., Wies, M., Kirwan, B., Arrigoni, V., Ruscio, D., Bos, T., Mohrmann, J.F.W., Maille, N., André, J.M., Letouzé, T., Diaz Pineda, J., Dormoy, C., Charles, R., Nixon J. (2018). D6.6: Test report large scale simulations. FutureSky Safety Joint Research Program.

Hruska, P., Hecker, K. G., Coderre, S., McLaughlin, K., Cortese, F., Doig, C., Beran, T., Wright, B., & Krigolson, O. (2016). Hemispheric activation differences in novice and expert clinicians during clinical decision making. *Advances in Health Sciences Education*, 21(5), 921–933.

Huijbrechts, E. J., & Van Paassen, M. M. (2023). Flight deck procedures for a new generation of pilots.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, 19(3), 265–288.

Iannello, P., Mottini, A., Tirelli, S., Riva, S., & Antonietti, A. (2017). Ambiguity and uncertainty tolerance, need for cognition, and their association with stress. A study among Italian practicing physicians. *Medical Education Online*, 22(1).

International Air Transport Association (IATA) (2002). Safety Report 2001.

International Air Transport Association (IATA). (2006). Safety report – 2005 edition.

International Air Transport Association (IATA). (2011). 2011 State of Global Aviation Safety.

International Air Transport Association (IATA). (2013). Safety report 2012.

International Air Transport Association (IATA) (2014a). Safety Report 2014.

International Air Transport Association (IATA). (2014b). Data report for evidence-based training (1st ed.).

International Air Transport Association (IATA). (2018). Safety report - 2018 edition.

International Air Transport Association (IATA) (2020). Safety Report 2019.

International Air Transport Association (IATA) (2024a). Competency-Based Training and Assessment (CBTA) Expansion within the Aviation System.

International Air Transport Association (IATA) (2024b). Evidence-Based Training Implementation Guide Edition 2.

International Air Transport Association (IATA). (2025a). 2024 Full Year Accident Update.

International Civil Aviation Organization (IATA). (2025b). State of global aviation safety – IATA Safety Report 2025 Edition.

International Air Transport Association (IATA). (2026a). IATA annual safety report 2025 (Executive Summary).

International Air Transport Association (IATA). (2026b). Annual review 2026.

International Air Transport Association (IATA). (2026c). Interactive safety report. Accessed July 7, 2026, https://www.iata.org/en/publications/safety-report/interactive-safety-report/

International Civil Aviation Organisation (ICAO). (2013). Doc 9995 Manual of Evidence-based Training, First Edition.

International Civil Aviation Organisation (ICAO). (2016). Doc 9966: Manual for the Oversight of Fatigue Management Approaches, Second Edition.

International Civil Aviation Organisation (ICAO). (2023). CBTA Instructional System Design.

International Civil Aviation Organization (ICAO). (2025). State of global aviation safety: 2025 edition (Safety Report).

Javaux, D., Fortmann, F., & Möhlenbrink, C. (2015). Adaptive human-automation cooperation: A general architecture for the flight deck and its application in the A-PiMod project. *Proceedings of the 7th International Conference on Advanced Cognitive Technologies and Applications*. International Academy, Research, and Industry Association (IARIA).

Julmi, C. (2019). When rational decision-making becomes irrational: a critical assessment and re-conceptualization of intuition effectiveness. *Business Research*, 12(1), 291–314.

Jung-Beeman, M. (2005). Bilateral brain processes for comprehending natural language. *Trends in Cognitive Sciences*, 9(11), 512–518.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, Daniel., Slovic, Paul., & Tversky, Amos. (1982). *Judgment under uncertainty: heuristics and biases*. Cambridge University Press.

Kelly, D., & Efthymiou, M. (2019). An analysis of human factors in fifty controlled flight into terrain aviation accidents from 2007 to 2017. *Journal of Safety Research*, *69*, 155–165.

Kent, P. (2017). Fluid intelligence: A brief history. *Applied Neuropsychology: Child, 6(3), 193-203*.

Kern, T. (2010). *Foundations of Professional Airmanship and Flight Discipline*. Convergent Performance. Colorado Springs, Colorado.

Kharoufah, H., Murray, J., Baxter, G., & Wild, G. (2018). A review of human factors causations in commercial air transport accidents and incidents: From to 2000–2016. *Progress in Aerospace Sciences*, *99*, 1–13.

Kirton, M. (1976). Adaptors and innovators: A description and measure. *Journal of applied psychology*, 61(5), 622.

Kirwan, B. (2025). Human factors requirements for human-AI teaming in aviation. *Future Transportation*, *5*(2), 42.

Klein, G. A. (2011). *Streetlights and shadows: Searching for the keys to adaptive decision making*. MIT Press.

Klein, G.A. (1996). Recognition-Primed Decision Strategies.

Klein, G.A. (1998) *Sources of Power: How People Make Decisions*. Cambridge, MA: MIT Press.

Klein, G.A. (2015). A naturalistic decision making perspective on studying intuitive decision making. *Journal of Applied Research in Memory and Cognition*, 4(3), 164–168.

Komite Nasional Keselamatan Transportasi (KNKT). (2015). Aircraft Accident Investigation Report PT. Indonesia Air Asia Airbus A320-216; PK-AXC.

Komite Nasional Keselamatan Transportasi (KNKT). (2019). Aircraft Accident Investigation Report PT. Lion Mentari Airlines Boeing 737-8 (MAX); PK-LQP.

Kounios, J., & Beeman, M. (2009). The Aha! moment: The cognitive neuroscience of insight. *Current Directions in Psychological Science*, 18(4), 210–216.

Kox, E., Rooij, J. V., Zon, R., & Kirwan, B. (2026). Human–AI Teaming in the Cockpit: From Theory to Practice—Views from the Flightdeck. *Aerospace*, *13*(10), 878.

Krain, A. L., Wilson, A. M., Arbuckle, R., Castellanos, F. X., & Milhama, M. P. (2006). Distinct neural mechanisms of risk and ambiguity: A meta-analysis of decision-making. *NeuroImage*, 32(1), 477–484.

Kühn, S., Ritter, S. M., Müller, B. C. N., van Baaren, R. B., Brass, M., & Dijksterhuis, A. (2014). The importance of the default mode network in creativity - A structural MRI study. *Journal of Creative Behavior*, 48(2), 152–163.

Kulida, E., & Lebedev, V. (2020). About the use of artificial intelligence methods in aviation. *2020 13th International Conference" Management of large-scale system development"(MLSD)* (pp. 1-5). IEEE.

Landman, A., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2017). Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise. *Human Factors*, *59*(8), 1161–1172.

Landman, A., van Middelaar, S. H., Groen, E. L., van Paassen, M. M., Bronkhorst, A. W., & Mulder, M. (2020). The effectiveness of a mnemonic-type startle and surprise management procedure for pilots. T*he International Journal of Aerospace Psychology*, 30(3-4), 104-118.

Landry, S.J. (2009). Flight Deck Automation. *Springer Handbook of Automation.* Springer Handbooks. Springer, Berlin, Heidelberg.

Langewiesche, W. (2009). *Fly by wire: The geese, the glide, the miracle on the Hudson.* Farrar, Straus and Giroux.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, 66, 799–823.

Leveson, N. G. (2016). *Engineering a safer world: Systems thinking applied to safety* (p. 560). The MIT Press.

Li, P. P. (2014). Toward the geocentric framework of intuition: The yin-yang balancing between the eastern and western perspectives on intuition. *Handbook of research methods on intuition* (pg. 28–41). Edward Elgar Publishing.

Li, W. C., Li, L. W., Harris, D., & Hsu, Y. L. (2014). The application of aeronautical decision-making support systems for improving pilots' performance in flight operations.

Li, W. C., Nichanian, A., Lin, J., & Braithwaite, G. (2024). What can we learn from severity index on flight data monitoring? Analysis of safety resilience in flight operations during COVID-19 disruptions. *Ergonomics*, 67(8), 1164-1176.

Li, Y., Chen, H., Xin, X., & Ji, M. (2020). The influence of mindfulness on mental state with regard to safety among civil pilots. *Journal of Air Transport Management*, 84.

Ligda, S. V., Fischer, U., Mosier, K., Matessa, M., Battiste, V., & Johnson, W. W. (2015). Effectiveness of advanced collaboration tools on crew communication in reduced crew operations. *International conference on engineering psychology and cognitive ergonomics* (pp. 416-427). Cham: Springer International Publishing.

Linstone, H. A., & Turoff, M. (1975). *The Delphi method* (Vol. 1975, pp. 3–12). Addison-Wesley.

Lintern, G., & Kugler, P. N. (2017). Sociotechnical System Safety: Hierarchical Control versus Mindfulness. *Systems Engineering*, 20(4), 307–317.

Lopes, N. M., Aparicio, M., Costa, C. J., Neves, F. T., & Bernardino, C. (2026). Designing the future of flight: A socio-technical framework for single-pilot operations in commercial aviation. *Journal of Air Transport Management*, 132, 102939.

Lorrig, P., & Daw, Z. (2024, September). Advances and Challenges Towards Enabling Human-AI-Teaming Applications for Flight Deck Operations. *2024 AIAA DATC/IEEE 43rd Digital Avionics Systems Conference (DASC)* (pp. 1-10). IEEE.

Loukopoulos, L.D., Dismukes, R.K., & Barshi, I. (2009). *The Multitasking Myth: Handling Complexity in Real-World Operations* (1st ed.). Routledge.

Low, A., & Purser, R. (2012). Zen and the creative management of dilemmas. *Journal of Management, Spirituality and Religion*, 9(4), 335–355.

Luchini, S. A., Volle, E., & Beaty, R. E. (2025). The role of the default mode network in creativity. *Current Opinion in Behavioral Sciences* (Vol. 65). Elsevier Ltd.

Malakis, S., Baumgartner, M., Berzina, N., Laursen, T., Smoker, A., Poti, A., & Kontogiannis, T. (2023). A framework for supporting adaptive human-AI teaming in air traffic control. International Conference on Human-Computer Interaction (pp. 320-330). Cham: Springer Nature Switzerland.

Malik, A., & Gollnick, V. (2016). Impact of reduced crew operations on airlines-Operational challenges and cost benefits. *16th AIAA aviation technology, integration, and operations conference* (p. 3303).

Malinge, Y. (2004). A300B4 loss of all hydraulics, Baghdad. *International Society of Air Safety Investigators (ISASI) Annual Seminar*.

Malinge, Y. (2011). The views from an aircraft manufacturer: What have we learned.

Man4Gen Consortium (2015). 6.7 - Final report of research methods and results.

Man4Gen Consortium. (2012). D1.1 literature review.

Man4Gen Consortium. (2015). Man4Gen Final Report Summary.

Manna, A., Raffone, A., Perrucci, M. G., Nardo, D., Ferretti, A., Tartaro, A., Londei, A., del Gratta, C., Belardinelli, M. O., & Romani, G. L. (2010). Neural correlates of focused attention and cognitive monitoring in meditation. *Brain Research Bulletin*, 82(1–2), 46–56.

Mansouri, F. A., Kievit, R. A., & Buckley, M. J. (2025). Executive control fluctuations underlie behavioral variability in anthropoids. *Trends in Cognitive Sciences* (Vol. 29, Issue 4, pp. 331–343). Elsevier Ltd.

Manzey, D., Reichenbach, J. & Onnasch L. (2012). Human performance consequences of automated decision aids: The impact of degree of automation and system experience. *Journal of Cognitive Engineering and Decision Making*.

Marchand, W. R. (2014). Neural mechanisms of mindfulness and meditation: Evidence from neuroimaging studies. *World Journal of Radiology*, 6(7), 471.

Martin, J. W., & Sloman, S. A. (2013). Refining the dual-system theory of choice. *Journal of Consumer Psychology*, 23(4), 552–555.

Martin, W., Murray, P., & Bates, P. (2012). The effects of startle on pilots during critical events: A case study analysis. *EAAP 30*. EAAP.

Matthews, G., Emo, A. K., Funke, G., Zeidner, M., & Roberts, R. D. (2003). Emotional intelligence: Implications for human factors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 47, No. 9, pp. 1053-1057). Sage CA: Los Angeles, CA: SAGE Publications.

Maxwell, G. A., & Grant, K. (2021). Commercial airline pilots' declining professional standing and increasing precarious employment. *The International Journal of Human Resource Management*, *32*(7), 1486-1508.

McCrea, S. M. (2010). Intuition, insight, and the right hemisphere: Emergence of higher sociocognitive functions. *Psychology Research and Behavior Management* (Vol. 3, pp. 1–39). Dove Medical Press Ltd.

McLain, D. L., Kefallonitis, E., & Armani, K. (2015). Ambiguity tolerance in organizations: Definitional clarification and perspectives on future research. *Frontiers in Psychology*, 6(MAR).

Meland, A., Ishimatsu, K., Pensgaard, A. M., Wagstaff, A., Fonne, V., Garde, A. H., & Harris, A. (2015). Impact of Mindfulness Training on Physiological Measures of Stress and Objective Measures of Attention Control in a Military Helicopter Unit. *International Journal of Aviation Psychology*, 25(3–4), 191–208.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function.*Brain structure & function* (Vol. 214, Issues 5–6, pp. 655–667).

Millar, M. (2012). Measuring Fatigue. *ICAO Asia-Pacific FRMS Seminar 2012*.

Miller, M. A. (2017). The Spire Mindfulness Tracker. *Alternative Medicine* (34), 12.

Minaskan, N., Alban-Dromoy, C., Pagani, A., Andre, J. M., & Stricker, D. (2022). Human intelligent machine teaming in single pilot operation: a case study. *International Conference on Human-Computer Interaction* (pp. 348-360). Cham: Springer International Publishing.

Mohrmann, J. F. W. (2013). Investigating flight crew recovery capabilities from system failures in highly automated fourth generation aircraft [Unpublished master's thesis]. Delft University of Technology.

Mohrmann, J.F.W. and Lemmers, A.J.J. (2013). Investigating Flight Crew Recovery Capabilities from System Failures in Highly Automated Fourth Generation Aircraft. NLR-TR-2013-574. National Aerospace Laboratory.

Mohrmann, J.F.W., Lemmers, A., & Stoop, J. (2015). Investigating flight crew recovery capabilities regarding system failures in highly automated fourth generation aircraft. *Aviation Psychology and Applied Human Factors*.

Mohrmann, J.F.W., Field, J.N. & Stoop, J. (2017). Resilient decision making in the flight deck: Does it work?. *Proceedings from Resilience Engineering Association Symposium Liége, Belgum, 26-29 june, 2017*. Resilience Engineering Association.

Mohrmann, J.F.W. & Stoop, J.A.A.M. (2019). Airmanship 2.0: Innovating aviation human factors forensics to necessarily proactive role. *International Society of Aviation Safety Investigators (ISASI) Annual Seminar*.

Mohrmann, J.F.W. & Field, J.N. (2026). Training for sensemaking in the modern flight deck: a new perspective [Preprint]

Mohrmann, J.F.W. (2026). Solving the brittleness problem: Redefining airmanship in the age of increasing complexity [Preprint]

Molen, P. P. van der (1990). The biological instability of social equilibria. *Sociobiology and conflict* (pp. 63-96). Springer, Dordrecht.

Molen, P.P. van der (1989) Adaption-Innovation and changes in the social structure: on the anatomy of catastrophe. In: Kirton, M.L. (Ed.) *Adaptors and Innovators: Styles of Creativity and Problem Solving*. Routledge, London and New York.

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.

Mumaw, R. J., Sarter, N., & Wickens, C. D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck. 11th International Symposium on Aviation Psychology*, Columbus, OH* (Vol. 8).

Myers III, P. L., & Starr Jr, A. W. (2021). Single pilot operations in commercial cockpits: background, challenges, and options. *Journal of Intelligent & Robotic Systems*, 102(1), 19.

Nadurak, V. (2023). Dual-Process Theory and Two Types of Metacognitive Monitoring and Control Processes. *Integrative Psychological and Behavioral Science*, 57(4), 1273–1283.

Naikar, N., Brady, A., Moy, G., & Kwok, H. W. (2023). Designing human-AI systems for complex settings: ideas from distributed, joint, and self-organising perspectives of sociotechnical systems and cognitive work analysis. *Ergonomics*, 66(11), 1669-1694.

National Transportation Safety Board (NTSB). (2014). Descent Below Visual Glidepath and Impact with Seawall, Asiana Airlines Flight 214, Boeing 777-200ER, HL7742, San Francisco, California, July 6, 2013. Aircraft Accident Report NTSB/AAR-14/01. Washington, DC.

Neisser, U. (1976). *Cognition and Reality*. W. H. Freemand.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, 104(9).

Niedermaier, D., Buch, J. P., Mohrmann, J.F.W., & Durak, U. (2018). Simulating the Unexpected: Challenge-Centric Simulator Scenario Design for Advanced Flight Crew Training. *2018 AIAA Modelling and Simulation Technologies Conference* (p. 1397).

Niedermeier, D., Papenfuß, A., & Wies, M. (2023). Simulator study on a spatially separated airline crew in a complex decision-making scenario. *AIAA AVIATION 2023 Forum* (p. 3474).

O'Doherty, J. P., Cockburn, J., & Pauli, W. M. (2017). Learning, Reward, and Decision Making. *Annual Review of Psychology* (Vol. 68, pp. 73–100). Annual Reviews Inc.

O'Hare, D. (1992). The 'artful' decision maker: A framework model for aeronautical decision making. *The international journal of aviation psychology*, *2*(3), 175-191.

Oken, B. S. (2016). A systems approach to stress and resilience in humans: Mindfulness meditation, aging, and cognitive function. Doctoral dissertation, Portland State University.

Oliver Wyman. (2025). Flight operations Q4 trends: Pilot shortage, costs, & Gen Z. https://www.oliverwyman.com/our-expertise/insights/2025/oct/flight-operations-trends-pilot-shortage-costs-gen-z.html

Öllinger, M., & Von Müller, A. (2017). Search and coherence-building in intuition and insight problem solving. *Frontiers in Psychology*, 8(MAY).

Orasanu, J., Martin, L., & Davison, J. (1998). Errors in aviation decision making: Bad decisions or bad luck?. National Aeronautics and Space Administration, Ames Research Center.

Oster, C. V., Strong, J. S., & Zorn, C. K. (2013). Why air travel is so safe: A review of recent aviation safety literature. *Research in Transportation Economics*, 43(1), 26-35.

Ovington, L. A., Saliba, A. J., Moran, C. C., Goldring, J., & MacDonald, J. B. (2018). Do People Really Have Insights in the Shower? The When, Where and Who of the Aha! Moment. *Journal of Creative Behavior*, 52(1), 21–34.

Pagnoni, G., Cekic, M., & Guo, Y. (2008). "Thinking about not-thinking": Neural correlates of conceptual processing during Zen meditation. *PLoS ONE*, 3(9).

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230–253.

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.

Pascal, B. (1995). *Pensées* (A. J. Krailsheimer, Trans.). Penguin Books. (Original work published 1670)

Pennycook, G. (2015). A perspective on the theoretical foundation of dual-process models. Psychology Press.

Poudel, R., Riedel, M. C., Salo, T., Flannery, J. S., Hill-Bowen, L. D., Eickhoff, S. B., Laird, A. R., & Sutherland, M. T. (2020). Common and distinct brain activity associated with risky and ambiguous decision-making. *Drug and Alcohol Dependence*, 209.

Prinzel, L. J., Krois, P., Ellis, K. K., Vincent, M., Stephens, C., Oza, N., ... & Matthews, B. L. (2024). The adaptable and resilient safety system: The human factor in future in-time aviation safety management systems. *AIAA SCITECH 2024 Forum* (p. 1603).

Proctor, R. W., & van Zandt, T. (2018). *Human Factors in Simple and Complex Systems,* Third Edition.

Rankin, A., Woltjer, R., Field, J. and Woods, D. (2013). "Staying ahead of the aircraft" and Managing Surprise in Modern Airliners. NLR-TP-2013-230. National Aerospace Laboratory.

Rankin, A., Ekström, E., Sjölin, V., Woltjer, R., Stepniczka, I., Eder, M. (2016). Man4Gen: Final report of research methods and results.

Rankin, A., Woltjer, R., & Field, J. (2016). Sensemaking following surprise in the flight deck—a re-framing problem. *Cognition, Technology & Work*, *18*(4), 623-642.

Rasmussen, J. (1983). Skills, Rules, and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-13(3), 257–266.

Reed, S. K. (2016). The Structure of Ill-Structured (and Well-Structured) Problems Revisited. *Educational Psychology Review* (Vol. 28, Issue 4, pp. 691–716). Springer New York LLC.

Reitsma, J. P. (2025). Elevated Levels of Automation for Aircraft System & Flight Plan Management. PhD Dissertation, Delft University of Technology.

Remmers, C., Topolinski, S., & Michalak, J. (2015). Mindful(l) intuition: Does mindfulness influence the access to intuitive processes? *Journal of Positive Psychology*, 10(3), 282–292.

Reuters. (2026). Can the US government be held liable for the LaGuardia Airport collision? *Reuters*. https://www.reuters.com/legal/litigation/can-us-government-be-held-liable-laguardia-airport-collision-2026-03-25/

Reyna, V. F. (2012). A new intuitionism: Meaning, memory, and development in fuzzy-trace theory. *Judgment and Decision Making*, 7(3), 332–359.

Rivera, J., Talone, A. B., Boesser, C. T., Jentsch, F., & Yeh, M. (2014). Startle and surprise on the flight deck : Similarities, differences, and prevalence. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 58(1), 1047-1051.

Rosenbloom, M. H., Schmahmann, J. D., & Price, B. H. (2012). The Functional Neuroanatomy of Decision-Making. *The Journal of Neuropsychiatry and Clinical Neurosciences* (Vol. 24).

Sadler-Smith, E., & Shefy, E. (2004). The intuitive executive: Understanding and applying "gut feel" in decision-making.

Salehinejad, M. A., Ghanavati, E., Rashid, M. H. A., & Nitsche, M. A. (2021). Hot and cold executive functions in the brain: A prefrontal-cingular network. *Brain and Neuroscience Advances* (Vol. 5). SAGE Publications Ltd.

Sallinen, M., van Dijk, H., Aeschbach, D., Maij, A., & Åkerstedt, T. (2020). A large-scale European Union study of aircrew fatigue during long night and disruptive duties. *Aerospace medicine and human performance, 91(8), 628-635.*

Sarter, N. B., & Woods, D. D. (1994). Pilot Interaction With Cockpit Automation 11: An Experimental Study of Pilots' Model and Awareness of the Flight Management System. *The International Journal of Aviation Psychology* (Vol. 4, Issue 1).

Sarter, N.B., Panesar, K., & Bhardwaj, A. (2023). The Contributions of Human Operators to Safety and Risk Mitigation: Implications for Crew Complements and Automation/Autonomy Levels in Commercial Transport Operations (No. DOT/FAA/TC-22/40). United States. Department of Transportation. Federal Aviation Administration. William J. Hughes Technical Center.

Sarter, N.B. (2024). Information Management on the Flight Deck of Highly Automated Aircraft.

Saurin, T. A., & Carim Junior, G. C. (2012). A framework for identifying and analyzing sources of resilience and brittleness: A case study of two air taxi carriers. *International Journal of Industrial Ergonomics*, *42*(3), 312–324.

Schmid, D., & Korn, B. (2017). A tripartite concept of a remote-copilot center for commercial single-pilot operations. *AIAA Information Systems-AIAA Infotech @ Aerospace* (p. 0064).

Schmid, D., & Stanton, N. A. (2020). Progressing toward airliners' reduced-crew operations: A systematic literature review. *The International Journal of Aerospace Psychology*, *30*(1-2), 1-24.

Schoemaker, P. J. (2004). Forecasting and scenario planning: The challenges of uncertainty and complexity. *Blackwell handbook of judgment and decision making*, 274-296.

Schutte, P. C. (2015). How to make the most of your human: Design considerations for single pilot operations. *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 480-491). Cham: Springer International Publishing.

Schwab, E. (2024). How Safe Is Safe Enough? Analyzing the Incentive Structure of the Products Liability Scheme on Commercial Aviation Manufacturers. U. Miami L. Rev., 79, 398.

Senate Committee (2020). US Senate Committee Investigation Report – Aviation Safety Oversight.

Shah, D., (2024). Timesaving: Why Airlines Operate Overnight Flights. *Simply Flying.* https://simpleflying.com/time-saving-airlines-overnight-flights/

Shappell, S. A., & Wiegmann, D. A. (2017). Applying reason: The human factors analysis and classification system (HFACS). *Human factors and aerospace safety* (pp. 59-86). Routledge.

Shaw, S. D., & Nave, G. (2026). Thinking-Fast, Slow, and Artificial: How AI is Reshaping Human Reasoning and the Rise of Cognitive Surrender.

Shively, R.J., Brandt, S. L., Lachter, J., Matessa, M., Sadler, G., & Battiste, H. (2016). Application of human-autonomy teaming (HAT) patterns to reduced crew operations (RCO). *International conference on engineering psychology and cognitive ergonomics* (pp. 244-255). Cham: Springer International Publishing.

Slagmolen, B., van Dalen, B., & Taen, R. (2009). *Mindful Organiseren.* Mediawerf Uitgevers Amsterdam.

Soo, K. (2018). The learning and development process of pilots during initial airline training.

Sprengart, S. M., Neis, S. M., & Schiefele, J. (2018). Role of the human operator in future commercial reduced crew operations. *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)* (pp. 1-10). IEEE.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate?

Stanovich, K. E. (2011). On the Distinction Between Rationality and Intelligence: Implications for Understanding Individual Diff erences in Reasoning.

Stanton, N. A., Harris, D., & Starr, A. (2016). The future flight deck: Modelling dual, single and distributed crewing options. *Applied Ergonomics*, *53*, 331-342.

Stoop, J. A., & van Kleef, E. A. (2015). Reliable or Resilient: Recovery from the Unanticipated. *International Journal of Performability Engineering* (Vol. 11, Issue 2).

Stoop, J. (2017). How did aviation become so safe, and beyond? *Proceedings of the 53rd ESReDA Seminar, 14 – 15 November 2017*. European Commission Joint Research Centre, Ispra, Italy

Stoycheva, K. (2010). Tolerance for ambiguity, creativity, and personality.

Stoycheva, K. (2025). Tolerance of Ambiguity and the Creative Action: To Engage and Endure. *Journal of Creative Behavior*, 59(3).

Strauch, B. (2017). *Investigating Human Error*. CRC Press.

Suijkerbuijk, H. C. H., Rouwhorst, W. F. J. A., Verhoeven, R. P. M., & Arents, R. R. D. (2017). Innovative flight deck touch screen HMI design using direct manipulation.

Svenson, F., Pietrzak, P., & Launer, M. A. (2023). Different types of intuition at the workplace: an integrative review. *Business: Theory and Practice*, 24(2), 334–348.

Swedish Accident Investigation Authority (SAIA). (2016). Final report RL2016:11e Accident in Oajevágge, Norrbotten County, Sweden on 8 January 2016 involving the aeroplane SE-DUX.

Talbot, M. (2004). Good wine may need to mature: A critique of accelerated higher specialist training. Evidence from cognitive neuroscience. *Medical Education* (Vol. 38, Issue 4, pp. 399–408).

Taleb, N. (2015). *The black swan: The impact of the highly improbable*. Victoria, 250, 595-7955.

Tang, Y. Y., Hölzel, B. K., & Posner, M. I. (2015). The neuroscience of mindfulness meditation. *Nature Reviews Neuroscience* (Vol. 16, Issue 4, pp. 213–225). Nature Publishing Group.

Ten Houten, W. D. (2016). Embodied Feeling and Reason in Decision-Making: Assessing the Somatic-Marker Hypothesis. *Sentimientos y razón corporeizados en la toma de decisiones: evaluando la hipótesis de los marcadores somáticos*. 20, 87–97.

Thompson, V. A. (2009). Dual-process theories: A metacognitive perspective.

Thompson, V.A., & Morsanyi, K. (2012). Analytic thinking: Do you feel like it? *Mind and Society*, 11(1), 93–105.

Tokadlı, G., Dorneich, M. C., & Matessa, M. (2021). Toward human–autonomy teaming in single-pilot operations: Domain analysis and requirements. *Journal of Air Transportation*, 29(4), 142-152.

Troadec, J.P. (2012). History of the Air France Flight 447 Accident. *ISASI Forum January-March 2013*.

UK Civil Aviation Authority (UK CAA). (2019). Pilot Fatigue Measurement Research Report.

Valenta, V. (2018). Effects of airline industry growth on pilot training. *MAD-Magazine of Aviation Development, 6(4), 52-56*.

Van den Berg, B., de Bruin, A. B. H., Marsman, J. B. C., Lorist, M. M., Schmidt, H. G., Aleman, A., & Snoek, J. W. (2020). Thinking fast or slow? Functional magnetic resonance imaging reveals stronger connectivity when experienced neurologists diagnose ambiguous cases. *Brain Communications*, 2(1).

Van Leeuwen, J. K. V., Landman, A., Groen, E. L., Mumaw, R. J., Stroosma, O., Paassen, M. M. V., & Mulder, M. (2024). Using problem-based exploratory training to improve pilot understanding of autopilot functions. *Cognition, Technology & Work*, 26(2), 267-279.

Vincenti, W., (1990). *What Engineers Know and How They Know It. Analytical Studies from Aeronautical History*. The John Hopkins University Press.

Vlaskamp, D., Landman, A., van Rooij, J., & Blundell, J. (2025). Recovery from startle and surprise: A survey of airline pilots' operational experience using a startle and surprise management method. *International Journal of Industrial Ergonomics*, 107, 103733.

Volz, K. G., Rübsamen, R., & von Cramon, D. Y. (2008). Cortical regions activated by the subjective sense of perceptual coherence of environmental sounds: A proposal for a neuroscience of intuition. *Cognitive, Affective and Behavioral Neuroscience*, 8(3), 318–328.

Volz, K. G., Schooler, L. J., & von Cramon, D. Y. (2010). It just felt right: The neural correlates of the fluency heuristic. *Consciousness and Cognition*, 19(3), 829–837.

Weick, K. E., & Sutcliffe, K. M. (2011). *Managing the unexpected: Resilient performance in an age of uncertainty* (Vol. 8). John Wiley & Sons.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics*, 23(10), 995–1011.

Wikander, R., & Dahlström, N. (2016). The multi crew pilot licence-revolution, evolution or not even a solution. *A review and analysis of the emergence, current situation and future of the multi-crew pilot licence (MPL)*. Lund University.

Williams, M., Palace, M., Welsh, JC., & Brooks, S. (2025). Neural correlates of chess expertise: A systematic review of brain imaging studies comparing expert versus novice players. *Brain Mechanisms*, 148–150, 202516.

Woiceshyn, J. (2022). *Handbook of intuition research as practice*. Edward Elgar Publishing.

Woltjer, R., Field, J., & Rankin, A. (2015). Adapting to the unexpected in the flight deck. In *Proceedings of the 6th resilience engineering association symposium*. Lisbon, Portugal. REA.

Woltjer, R., Rankin, A., Ekström, E., Sjölin, V., & Field, J. (2026). Investigating flight crew strategies to cope with unexpected events: a multi-layered extended control model of joint crew-automation activity. *Safety Science*, *199*, 107191.

Wood, S. (2004). Flight crew reliance on automation. Civil Aviation Authority.

Woods, D. D. (2015). Four concepts for resilience and the implications for the future of resilience engineering. *Reliability engineering & system safety, 141, 5-9*.

Woods, D. D. (2019). Essentials of resilience, revisited. *Handbook on resilience of socio-technical systems*, 52-65.

World Bank. (2026). Air transport, registered carrier departures worldwide. World Development Indicators. Retrieved July 7, 2026, from https://data.worldbank.org/indicator/IS.AIR.DPRT

Zamani, M. R. (2019). Neural Areas of Activation During Feedback Processing of Clinical Decision Making. Doctoral dissertation.

Zamarian, L., Sinz, H., Bonatti, E., Gamboz, N., & Delazer, M. (2008). Normal Aging Affects Decisions Under Ambiguity, but Not Decisions Under Risk. *Neuropsychology*, 22(5), 645–657.

Zander, T., Öllinger, M., & Volz, K. G. (2016). Intuition and insight: Two processes that build on each other or fundamentally differ? *Frontiers in Psychology*, 7(SEP).

Zedelius, C. M., & Schooler, J. W. (2015). Mind wandering "Ahas" versus mindful reasoning: alternative routes to creative solutions. *Frontiers in Psychology*, 6.

Zeki, S. (2004). The neurology of ambiguity. *Consciousness and Cognition*, 13(1), 173–196.

Ziegler, M., Danay, E., Heene, M., Asendorpf, J., & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality, 46(2), 173-183*.

Zou, Y. (2025). Transparent Path Planning For Uncrewed Air Traffic Management. Doctoral Dissertation. Delft University of Technology.

Abbreviations

AA	Automation Awareness
ACC	Anterior Cingular Cortex
AI	Artificial Intelligence
ANCOVA	Analysis of covariance
AOA	Angle of Attack
ATC	Air Traffic Control
ATM	Air Traffic Management
ATPL	Air Transport Pilot License
AVES	Air Vehicle Simulator
BH	Benjamini-Hochberg
BSL	Baseline group
CAST	Causal Analysis Using System Theory
CBTA	Competency Based Training & Assessment
CM	Contingency Management
COCOM	Contextual Control Model
CPDLC	Controller Pilot Data Link Communications
CQ	Communication Quality
CRM	Crew Resource Management
CVR	Cockpit Voice Recorder
dACC	dorsal Anterior Cingular Cortex
DESIDE	Detect, Estimate, Set safety objectives, Identify, Do, Evaluate
DFCP	Desired Flight Crew Performance
DIS	Display group (with RIS)
dlPFC	dorso-lateral Pre-Frontal Cortex
DLR	German Aerospace Centre
DMN	Default Mode Network
DPT	Dual Process Theory
EASA	European Union Aviation Safety Agency
EBT	Evidence Based Training

ECAM	Electronic Centralized Aircraft Monitor
ECOM	Extended Contextual Control Model
EF	Executive Function
EI	Emotional Intelligence
EICAS	Engine Indication and Crew Alerting System
eMCO	enhanced Minimum Crew Operations
ENG	Engine
ETOPS	Extended Twin Engine Operations
EU	European Union
F/O	First Officer
FAA	Federal Aviation Administration
FAC	Flight Augmentation Computers
FADEC	Full Authority Digital Engine Control
FCDC	Flight Control Data Concentrator
FCMC	Flight Control and Management Computer
FCOM	Flight Crew Operating Manual
FDR	False Discovery Rate
FDSTS	Flight Deck Socio-Technical System
FL	Flight Level
FLCH	Flight Level Change
FMS	Flight Management System
FOA	Fuel on Arrival
FORDEC	Facts, Options, Risks and benefits, Decide, Execute, Check
FRMS	Fatigue Risk Management System
FSF	Flight Safety Foundation
GEN	Generator
GPS	Global Positioning System
GEN	Generator
HAT	Human-Artificial Intelligence Teaming
HFACS	Human Factors Analysis and Classification System
HMI	Human Machine Interface
HRO	High Reliability Organisation

IATA	International Air Transport Association
ICAO	International Civil Aviation Organisation
IDG	Integrated Drive Generator
IEC	Instructor & Examiner Competency
JCS	Joint Cognitive System
KW	Kruskal-Wallis
LIFUS	Line Flying Under Supervision
LOC-I	Loss of Control - In Flight
LOFT	Line Oriented Flight Training
(M)MEL	(Master) Minimum Equipment List
Man4Gen	Manual Operations of Fourth Generation Airlines
MCAS	Maneuvering Characteristics Augmentation System
MCC	Multi-Crew Coordination
MCT	Max Continuous Thrust
MPL	Multi-Pilot License
NLR	Royal Netherlands Aerospace Centre
NM	Nautical Mile
OEM	Original Equipment Manufacturer
OFC	Orbito-Frontal Cortex
OLS	Ordinary Least Squares
ORCA	Observe Record Classify Assess
PC1-3	Principle Characteristic 1-3
PCC	Posterior Cingular Cortex
PCM	Perceptual Control Cycle
PFD	Primary Flight Display
QRH	Quick Reference Handbook
RA	Recursive Abstraction
RCO	Reduced Crew Operations
RIS	Risk Information System
RPD	Recognition Primed Decision Making
RA	Recursive Abstraction
S/O	Second Officer

SA	Situation Awareness
SC1-5	Strategy Characteristic 1-5
SFCC	Slat Flap Control Computer
SM	Sensemaking
SOP	Standard Operation Procedures
SPO	Single Pilot Operations
SRK	Skills Rules Knowledge
STAMP	System Theoretic Accident Model and Processes
STG	Strategy group
STPA	Systems-Theoretic Process Analysis
(T-)DODAR	(Time,) Diagnose, Options, Decide, Assign, and Review
TEM	Theat and Error Management
TM	Time Management
UK CAA	United Kingdom Civil Aviation Authority
UM	Uncertainty Management
UPRT	Upset Prevention and Recovery Training
USD	United States Dollars
vACC	ventral Anterior Cingular Cortex
VIF	Variable Inflation Factor
vlPFC	ventro-lateral Pre-Frontal Cortex
vmPFC	ventro-medial Pre-Frontal Cortex
WLB	Work Life Balance
YOY	Year over Year
ZFTT	Zero Flight Time Training

List of publications

Mohrmann, J.F.W. (2026). Why we need more showers in the aircraft: A neurocognitive perspective on the future pilot [Preprint engrXiv]

Mohrmann, J.F.W. (2026). Solving the brittleness problem: Redefining airmanship in the age of increasing complexity [Preprint engrXiv]

Mohrmann, J.F.W. & Field, J.N. (2026). Training for sensemaking in the modern flight deck: a new perspective [Preprint engrXiv]

Mohrmann, J.F.W. & Stoop, J.A.A.M. (2019). Airmanship 2.0: Innovating aviation human factors forensics to necessarily proactive role. *International Society of Aviation Safety Investigators (ISASI) Annual Seminar*.

Niedermeier, D., Buch, J. P., Mohrmann, J.F.W., & Durak, U. (2018). Simulating the Unexpected: Challenge-Centric Simulator Scenario Design for Advanced Flight Crew Training. *2018 AIAA Modeling and Simulation Technologies Conference* (p. 1397).

Field, J.N., Gerretsen, A., Meiland, R., & Mohrmann, J.F.W. (2017). Harmonizing the Scenario Generation Process, Tools and Standards for Flight Simulators. *AIAA Modeling and Simulation Technologies Conference* (p. 1310).

Mohrmann, J.F.W., Field, J.N. & Stoop, J.A.A.M. (2017). Resilient decision making in the flight deck: Does it work?. *Proceedings from Resilience Engineering Association Symposium Liége, Belgum, 26-29 june, 2017*. Resilience Engineering Association.

Field, J.N., Rankin, A., Mohrmann, J.F.W., Boland, E., & Woltjer, R. (2017). Flexible procedures to deal with complex unexpected events in the cockpit. *Resilience Engineering Association Symposium Liége, Belgum, 26-29 june, 2017*. Resilience Engineering Association.

Field, J.N., Fucke, L., Correia Grácio, B., & Mohrmann, J.F.W. (2016). Flight crew response to unexpected events: a simulator experiment. *AIAA Modeling and Simulation Technologies Conference* (p. 3373).

Mohrmann, J.F.W., Lemmers, A., & Stoop, J.A.A.M. (2015). Investigating flight crew recovery capabilities regarding system failures in highly automated fourth generation aircraft. *Aviation Psychology and Applied Human Factors*.

Mohrmann, J. F. W. (2013). Investigating flight crew recovery capabilities from system failures in highly automated fourth generation aircraft [Unpublished master's thesis]. Delft University of Technology.

Curriculum vitae

Frederik Mohrmann was born on March 21, 1988 in Amsterdam, The Netherlands. He was raised abroad across several continents, completing the International Baccalaureate in Washington D.C. in 2006, after which he moved to Delft to study aerospace. In 2010 he received his BSc in Aerospace Engineering. During his master's degree studies, he was acquainted and inspired by forensic engineering and accident investigation by early academic guide Prof.dr.ir. John Stoop. This inspired an internship at the National Transportation Safety Board in Washington D.C. and a desire to better understand the existing challenges in human-machine dynamics in modern cockpits. His MSc thesis was performed at the Royal Netherlands Aerospace Centre (NLR), orchestrating a large-scale simulator study as part of the Man4Gen EU program, titled "Investigating flight crew recovery capabilities regarding system failures in highly automated fourth generation aircraft.". He completed his MSc in Aerospace Engineering in 2013, his research winning best paper award in the Journal of Aviation Psychology and Applied Human Factors in 2015.

After his studies, Frederik continued with applied research in the flight deck and aircrew performance domain at NLR, where he was engaged in several large EU research consortia investigating new training and automation support concepts. With a desire to have greater positive impact on training standards, he spearheaded NLR's support of EBT and CBTA in the industry, establishing as an independent advisor to EASA in rulemaking for performance-based training and FSTD characteristics signature regulations. He has supported multiple operators and countries in the adoption of CBTA and EBT practices. Through this, he established NLR's innovation management program to improve, supporting research outcomes and ideas in their successful evolution and transition to impactful and market-validated products and services across all aerospace divisions.

In 2021 formally registered at the Delft Aerospace Faculty as an external PhD candidate, returning to the Department of Control & Simulation under the guidance of Prof.dr.ir. Max Mulder and Prof.dr. Sidney Dekker, whose mutual expertise fitted the subject of Airmanship 2.0. In this period, he also served at Boeing as a global representative for evolving aircrew training standards and regulations, supporting Boeing aircraft operators in transitioning to EBT and CBTA globally, as well as leading Boeing's portfolio strategy for data-driven training and support programs.

As a recognized global expert in the domain of advanced aircrew training, in 2026 he established the advisory firm Areti One together with his long-time training companion Captain Alex Wells, to support commercial air operators, aviation training organizations and authorities in evolving aircrew training toward CBTA and EBT, integrating innovations in training and simulation techniques. In his spare time, he practices what he preaches as a private pilot, glider instructor and aerobatics trainer, inspiring higher performance and airmanship in the next generation of aviators.

Propositions

Accompanying the dissertation *Airmanship 2.0 From Human Error to Human Asset* by Johannes Frederik Willem Mohrmann.

1. *Flight safety positively relates to flight crews' cognitive flexibility and adaptability when facing opacity and ambiguity in the flight deck. (Chapter 2)*

2. *Creating time for diagnostics and engaging in a cyclic process of sensemaking leads to more effective management of opacity. (Chapter 2 & 3)*

3. *Deconflicting the current pilot directives of 'be compliant' and 'be creative' by nesting the former in the latter facilitates a broader repertoire of neurocognitive sensemaking abilities. (Chapter 4)*

4. *Human sensemaking to manage opacity and ambiguity is more effective when the analytical and intuitive neurocognitive pathways are properly leveraged. (Chapter 5)*

5. *Existing human factors frameworks do not feature the resolution required to model behaviors in flight operations that feature complexity, opacity and ambiguity.*

6. *Greater complexity and autonomy in aviation systems and operations require increased utility of the human asset, not less.*

7. *The introduction of neurocognitive sensemaking training in commercial aviation will be uncomfortable for and elicit opposition from pilots, engineers and regulators that still clutch to the illusion of determinism.*

8. *Making sense of sensemaking at an aerospace engineering faculty requires institutional tolerance of ambiguity.*

9. *Sometimes locking a PhD candidate in a writing cottage is in the interest of a timely completion.*

10. *Seeing ourselves as liabilities as Taylor did[29], brought us good performance at the cost of great performance.*

These propositions are regarded as opposable and defendable, and have been approved as such by the promotors Prof.dr.ir. M. Mulder and Prof.dr. S.W.A. Dekker.

[29] Reference to Taylor, F.W. (1911). *The principles of scientific management.* Harper & Brothers.

After decades of enhancing safety through standardization, procedures and compliance, the aviation industry is reaching the limits of that safety paradigm. High operational reliability, reliance on compliance and emergent behavior of complex systems induce a characteristic brittleness in the flight deck socio-technical system, leading to critical shortfalls in the face of opaque conditions.

To access the next echelon of safety performance, the human pilot may yet provide the antidotal capacity to learn and adapt in the face of opacity and complexity. In making sense of sensemaking, this journey provides a fresh perspective on human aviation by re-examining what it takes to be an aviator in the 21st century.

