CRANFIELD UNIVERSITY

SQUADRON LEADER RICHARD GREEN MENG
(HONS) BSC (HONS) CENG MIET RAF

SUPPORTING OPERATIONAL DECISION MAKING CONCERNING AIRCRAFT STRUCTURAL INTEGRITY DAMAGE IDENTIFIED DURING MAINTENANCE.

SCHOOL OF DEFENCE AND SECURITY
Centre for Simulation and Analytics

PhD
Academic Year: 2020–2021

Supervisors: Dr K. R. McNaught, Dr A. Saddington
10 June 2021

CRANFIELD UNIVERSITY

SCHOOL OF DEFENCE AND SECURITY
Centre for Simulation and Analytics

PhD

Academic Year: 2020–2021

SQUADRON LEADER RICHARD GREEN MENG
(HONS) BSC (HONS) CENG MIET RAF

Supporting operational decision making concerning aircraft structural integrity damage identified during maintenance.

Supervisors: Dr K. R. McNaught, Dr A. Saddington
10 June 2021

© Crown Copyright 2021. All rights reserved. No part of this publication may be reproduced without the written permission of the copyright owner.

Abstract

Military aircraft operations balance delivery pressures and engineering risks. Aircraft structural damage incurred in-service creates complex risk decision problems for managers deliberating maintenance activity such as delaying rectification to continue operations, or grounding an aircraft or entire fleet. In many operational settings, aircraft availability demands restrict the time, information, or resources to analyse structural risks, making formal risk or decision analysis intractable. Exact solutions are information intensive and require specialist knowledge or machinery beyond the capabilities of generalist engineering managers, often compelling decision-makers to use their subjective judgement in an unsupported way. For actors deliberating aircraft maintenance structural risks in such circumstances, a novel approach based upon heuristics, argument and bounded rationality is proposed, which was informed by the results from a survey of engineering practitioners and case study analyses. Testing of the approach was carried-out with 21 aircraft engineering decision-makers with experience of structural integrity risks, split into three groups, using realistic but fictional textual simulations of aircraft maintenance. One group used existing decision justification approaches and were compared with a second group who provided decision justifications using the novel approach. Users of the novel approach felt supported and were very confident in their justifications. The third group of raters comparing the two sets of decision justifications indicated preferences using Likert scales against the criteria: which is easier to understand, which is more transparent, and which gives the better justification. Analysis of the comparative results

using ANOVA provided evidence that the novel approach enabled better decision justification and transparency compared to existing approaches. The novel approach aids decision-makers compelled to use their unsupported subjective judgement, improving organisational resilience by improving robustness and stretching system process to handle surprises, and providing a clear record of the decision basis for post hoc review.

The research project was funded by the UK Ministry of Defence, Royal Air Force Engineer Branch.

Keywords

Aircraft maintenance; structural integrity; risk assessment; naturalistic decision-making; argument; bounded rationality.

Acknowledgements

My PhD research journey has been one of the most challenging projects I have undertaken mentally, emotionally and academically. Reaching the finish line mostly unscathed could not have been possible without the advocacy and support of a number of key people, who I would like recognised within this thesis. Firstly, as the project was sponsored by the Royal Air Force Engineer Branch, supporters, colleagues and mentors enabled the research project to proceed in the first place: Dave Tozer, Jase Hill, Stu Simpson, Howard Fielder, Manjeet Ghatora, Paul Regan, and Andy Lewis. My supervisors and colleagues at Cranfield University have both steered and scrutinised the project, and been a reassuring shoulder particularly during difficult periods: Dr Ken McNaught, Dr Alistair Saddington, Prof Kevin Knowles, Laura Lacey, Bea Kingdon, Mandy Smith, and Clare Knock. Although I am unable to name them, I am extremely grateful to the busy professionals who gave their time as either interviewees, survey participants, pilot testers, framework testers, raters and control participants, or for ad hoc discussion. Lastly, and most importantly, I am indebted to my family and friends who have encouraged and inspired me through the most challenging periods: my wife Rhea, son James, Dad, Mum, brother Mike, and Eric.

Contents

Abstract iii

Contents vi

List of Figures ix

List of Tables xii

List of Abbreviations xiv

1 Introduction 1
- 1.1 Research context . . . 5
- 1.2 Predicting aircraft structural failure . . . 8
- 1.3 Research aim and objectives . . . 14
- 1.4 Overview of research design and methodology . . . 15
- 1.5 Thesis layout . . . 17
- 1.6 Chapter summary . . . 19

2 Literature review 20
- 2.1 Aviation safety . . . 21
- 2.2 Risk domain review . . . 23
- 2.3 Decision analysis domain review . . . 63
- 2.4 Literature review findings . . . 83
- 2.5 Chapter summary . . . 85

3 Initial data collection methodology 87
- 3.1 Initial data capture and analysis (Phase 2) . . . 88
- 3.2 Survey data collection design . . . 89
- 3.3 Case study data collection design . . . 100
- 3.4 Chapter summary . . . 101

4 Survey and case study results and discussion 104
- 4.1 Practitioner survey results . . . 104
- 4.2 Phase 2 - case study results . . . 148
- 4.3 Survey and case study discussion . . . 162
- 4.4 Chapter summary . . . 189

5 Novel framework methodology 190
5.1 Novel framework development (Phase 3) . . . 191
5.2 Novel framework testing methodology (Phase 4) . . . 220
5.3 Chapter summary . . . 244

6 Novel framework test results and discussion 245
6.1 Results from the first phase of the safety decision experiment . . . 246
6.2 Results from the second phase of the safety decision experiment . 266
6.3 Novel framework testing discussion . . . 271
6.4 Chapter summary . . . 281

7 Conclusions 283
7.1 Satisfaction of aim and objectives . . . 284
7.2 Contributions to knowledge . . . 288
7.3 Implications and practical significance . . . 289
7.4 Limitations and future work . . . 291
7.5 Final thoughts . . . 293

References 296

A Project Gantt chart 315

B Participant recruitment marketing 316

C Interview transcripts 318

D Main survey questions 403

E Main survey screenshots 410

F Submitted journal article 413

G Thematic analysis coding and sorting photographs 449

H Outstanding survey responses 451

I Framework development requirements 455
I.1 User requirements . . . 455
I.2 System requirements trace . . . 460

J Framework testing recruitment 463
J.1 Practitioner survey (Phase 2) participant recruitment invitation . . . 463
J.2 Framework testing (Phase 4) participant recruitment invitations . . 464

K Framework testing scenario information 467
K.1 Phase 4.1 participant context and case study information . . . 468
K.2 Phase 4.2 participant context information . . . 481
K.3 Framework testing participant questions . . . 485

L Novel framework training manuals **487**
L.1 Full Training Manual 487
L.2 Quick Reference Guide 512

List of Figures

1.1 The simplified aircraft maintenance decision problem. 3
1.2 Aircraft maintenance structural integrity risk decision flow diagram. 5
1.3 Process flow of the research project. 17

2.1 Risk matrix example. 33
2.2 A methodology for risk analysis. 37
2.3 A classification of techniques for predicting structural failure. 51
2.4 Dynamic Bayesian Network representing a deteriorating structural model. 56
2.5 Hierarchy of Decisions. 64
2.6 An extended decision-making hierarchy. 75

4.1 Practitioner survey demographics. 105
4.2 Survey Question 6 qualitative analysis. 106
4.2 Survey Question 6 qualitative analysis (cont.) 107
4.3 Survey Question 7a thematic map. 108
4.4 Responses to Question 8. 109
4.5 Survey Question 8b thematic map. 110
4.6 Survey Question 8e thematic analysis. 112
4.7 Survey Question 9a responses thematic map. 113
4.8 Survey Question 9b responses plot, grouped by respondent categories and roles. 114
4.9 Survey Question 9c responses plot. 115
4.10 Survey Question 9e thematic analysis results. 119
4.11 Aggregated responses to survey Question 11 and 12 with confidence intervals. 120
4.12 Survey Question 13 and 13a results analysis. 121
4.13 Survey Question 14 and 14a results analysis. 122
4.14 Survey Question 15 and 15a results analysis. 124
4.15 Survey Question 16 and 16a results analysis. 126
4.16 Survey Question 18 thematic analysis results. 127
4.17 Survey Question 19 thematic analysis results. 129
4.18 Survey Question 20 responses. 130
4.19 Survey Question 21a responses. 132
4.20 Survey Question 21b2 thematic analysis results. 133
4.21 Survey Question 21c thematic analysis results. 134
4.22 Survey Question 21d thematic analysis results. 135

4.23 Survey Question 26a thematic analysis results. 139
4.24 Survey Question 34 thematic analysis results. 145
4.25 Survey Question 36 thematic analysis results. 147
4.26 Initial data capture discussion thematic map 163
4.27 A 'supported' Type 2 decision thought process. 186
4.28 An 'unsupported' Type 2 decision thought process. 186

5.1 Screen capture from novel framework development. 194
5.2 Novel framework heuristic decision tree. 205
5.3 The Toulmin schema. 215
5.4 The holistic decision-making situation. 217
5.5 F-15 general arrangement, original and amended for simulation. . 227
5.6 Initial sort of Phase 4 participants. 233
5.7 Sub-phase 4.1 experience pair sort. 234
5.8 Sub-phase 4.2 participant allocation plan. 235
5.9 Allocation of participants for Sub-phase 4.2 testing. 236
5.10 Intervention participant data input fields. 240
5.11 Phase 4.1 intervention participant testing screen captures. 241
5.12 Phase 4.1 control participant testing screen captures. 242
5.13 Phase 4.2 participant testing screen capture. 243

6.1 Intervention participant reflection responses regarding novel framework operational usability. 248
6.2 Participant SIRM-Phase4-2020_16 response to case 4. 248
6.3 Phase 4 intervention and control cohort framework feedback. . . . 249
6.4 Intervention cohort confidence levels for each test case. 250
6.5 Examples of intervention responses to Sub-phase 4.1. 252
6.6 Descriptive statistics of completeness variation across intervention participants. 254
6.7 Comparison of the top 20 *class pairings*. 259
6.8 Comparison of information case/decision class pairs, marginalised by *experience pairing*. 260
6.9 Comparison of case information classes used by participants. . . . 261
6.10 Comparison of information classes for *discriminating* case lines. . 262
6.11 Comparison of control and intervention cohort decisions across each test case. 263
6.12 Counts of Toulmin schema elements within control participant decision justifications. 265
6.13 Participant SIRM-Phase4-2020_16 response to case 3. 280

B.1 Participant handout flyer distributed at internal UK MoD Royal Air Force Engineering Branch conference, 2018. 316
B.2 Participant recruitment poster presented at internal UK MOD Royal Air Force Engineering Branch conference, 2018. 317

E.1 Practitioner survey screen capture no. 1. 410
E.2 Practitioner survey screen capture no. 2. 411

E.3 Practitioner survey screen capture no. 3. 412

G.1 Survey results thematic analysis typical collection of themes, elements or factors prior to sorting and grouping. 449
G.2 Survey results thematic analysis typical collection of themes following grouping into codes. 450

H.1 Survey Questions 29b and 29c thematic analysis results. 453

List of Tables

1.1 Structural failure mechanisms in engineering and aircraft components. 9

2.1 Descriptions of normative and descriptive decision analysis. 67

3.1 Codes used for case studies. 102

4.1 Top 3 risk analysis processes used to support structural integrity risk analysis and decision-making. 116
4.2 Participant differing preferences regarding risk representation. . . 174
4.3 Prevalence of inconsistencies in survey responses by question. . . 178
4.4 Respondent inconsistencies during survey by role and category. . 178

5.1 Themes and sub-themes within user requirements. 195
5.1 (continued) 196
5.2 Expressions used to codify the case studies in identifying unsupported decision judgements to risky situations. 213
5.3 Information available to participants in phase 4 test cases. 229

6.1 Comparison of control and intervention cohort information use. . . 255
6.2 Case information class codes. 257
6.3 Decision class codes. 258
6.4 Differences in case information classes across control and intervention. 262
6.5 Intervention cohort correlations between participant and alternate choice. 264
6.6 ANOVA table for the response variable “ease of understanding”, given the factors *RaterID*, *Case* and *Pairing*. 269
6.7 ANOVA table for the response variable “greater transparency”, given the factors *RaterID*, *Case* and *Pairing*. 270
6.8 ANOVA table for the response variable “better justification”, given the factors *RaterID*, *Case* and *Pairing*. 270
6.9 Descriptive statistics for the three response variables, arranged by the completeness level of the intervention assessment. 271

I.1 User requirements generated from the Phase 2 survey results. . . 455
I.1 (continued) 456
I.1 (continued) 457
I.1 (continued) 458

I.1 (continued) 459

List of Abbreviations

AD	Airworthiness directive
ADF	Acceptable deferred fault
ALARP	As low as reasonably practicable
ANOVA	Analysis of variance
ARS	Acceptability, relevance, sufficiency argument evaluative criteria
ASIP	Aircraft Structural Integrity Program
ASSC	Air system safety case
BMAR	Baseline military airworthiness review
BN	Bayesian network
CAA	Civil Aviation Authority
CBA	Cost benefit analysis
CURES	Cranfield University Research Ethics System
DAOS	Design approved organisation system
DBN	Dynamic Bayesian network
DefStan	UK Defence Standard
DH	Duty holder
DHAN	Duty holder advice note
DO	Design organisation
DT	Delivery Team
EA	UK MOD aviation Engineering Authority
EASA	European Union Aviation Safety Agency

EIFS	Equivalent initial flaw size
ESVRE	Establishing, sustaining, validating, recovering, exploiting SI
FAA	Federal Aviation Authority
FFH	Fast and frugal heuristics
FH	Flying hour
FLE	Fatigue life expended
FMEA	Failure mode and effects analysis
FSR	Field service representative
FTA	Fault tree analysis
HAZOP	Hazard and operability method
HEA	Human error analysis
HEI	Human error identification
HRA	Human reliability analysis
HSE	UK Health and Safety Executive
ISAA	Independent Structural Airworthiness Advisor
LOAA	Letter of airworthiness authority
MAA	Military Aviation Authority
MAR	Military airworthiness review
MTI	Morton Thiokol
MOD	UK Ministry of Defence
MODREC	MOD Research and Ethics Committee
NASA	National Aeronautics and Space Administration
NDM	Naturalistic decision-making
NDT/I	Non-destructive testing/inspection
OEM	Original equipment manufacturer
ODH	Operational Duty Holder
ORR	Office of Rail and Road
PDT	Post design task

PRA	Probabilistic risk analysis
PRAT	Proportional risk assessment technique
PT	Project Team
QA	Quality assurance/audit
QRA	Quantitative risk analysis
RA	MAA regulatory article
RCM	Reliability centred maintenance
RtL	Risk to life
RUL	Remaining useful life
SEIN	Significant engineering information notice
SEU	Subjective expected utility
SHM	Structural health monitoring
SI	Structural integrity
SIWG	Structural integrity working group
SME	Subject matter expert
SQEP	Suitably qualified and experienced personnel
SRB	Solid rocket booster
SRM	Solid rocket motor
SSI	Structurally significant item
STS	Shuttle transport system
TAA	Type Airworthiness Authority
USAF	United States Air Force

Chapter 1

Introduction

"*. . . This is [also] a tendency of decision-makers and agencies who prefer not to talk about the uncertainties underlying the risk values assessed, acknowledging uncertainty can weaken the authority of the decision-maker and agency by creating an image of being unknowledgeable. Precise numbers are preferably used as a facade to cover up what are often political decisions.*"
(Aven and Zio, 2014)

"*. . . No matter how long you have studied the subject, solving probability problems intuitively is as likely to be successful as a pilot flying in bad weather without visible reference and without instruments. The list of people who have made reasoning errors about uncertainty looks like the roll call of famous scientists*" (Howard and Abbas, 2016)

Aircraft maintenance decision-makers are regularly faced with situations that necessitate balancing the requirements of airworthiness, equipment availability, and maintenance efficiency. When faults and damage to an aircraft are identified during the operational phase of an aircraft's life cycle, appropriately authorised and responsible maintenance engineers may permit the aircraft to continue operating with the fault unrectified with restrictions to operating capability, duration, mission or payload, or with no restriction. The alternative is cease operations until the fault is rectified or investigated further. These decisions necessarily consider the objective risk presented by a fault to aircraft safety and operational capacity in addition to the organisational values and wider societal considerations, such as company and personal reputation, environmental impact of a crash, and risk to those being overflown.

"Risk" is *"the effect of uncertainty on objectives"*(The British Standards Institution, 2018a). The objectives affected could be a loss of money, life, or other consequences of value to people (Howard and Abbas, 2016) and uncertainty could regard the chances of event occurrence, the effect on the objectives or the objectives affected (Yates and Stone, 1992, 11-15).

Aircraft operations are heavily regulated. The regulations prescribe safe practices to assure airworthiness[1] and reduce the risk to human life and injury to reasonably practicable levels. Regulations provide pre-conceived rule-based mitigation of uncertainty over the whole life cycle of an aircraft from design to disposal (for example, specifying safety factors to account for scatter fatigue strength of structure). In particular, some regulations control behaviours by either setting threshold risk levels[2], or by necessitating that risks are demonstrably controlled[3]. Regulators control and monitor organisational risk management policies for aircraft maintenance to assure governmental and public stakeholders. Standardising risk measures simplifies the communication of the applicability and standards required. However, in maintenance situations where there is deliberation over structural damage and faults, the pre-defined rules create tension in decision-makers expected to meaningfully balance risks with aircraft availability requirements.

Simplification of the maintenance decision problem concentrates on the immediate decision and its outputs as shown in Figure 1.1. Here the value p

[1] Airworthiness is defined as "The ability of an Air System or other airborne equipment or system to be operated in flight and on the ground without significant hazard to aircrew, ground crew, passengers or to third parties; it is a technical attribute of materiel throughout its lifecycle." (Military Aviation Authority, 2020a)

[2] European Union Aviation Safety Agency regulations require operating risk to a quantified average of less than one catastrophic event in 10 million flying hours (European Aviation Safety Agency, 2018, p46) that aircraft operators must demonstrate that they operate within. The UK Ministry of Defence (MoD) aviation regulator, the Military Aviation Authority (MAA) requires the design of air systems to have a cumulative probability of loss of an aircraft due to technical failure (inclusive of systems) assessed as no more frequent that one in one million flying hours (Military Aviation Authority, 2018b).

[3] The MAA requires risks to life to be As Low As Reasonably Practicable (ALARP) and tolerable. These measures necessitate consideration of the situation, including the extent to which risk reduction measures are grossly disproportionate (in time, money and trouble) to implement against the expected safety benefit (Military Aviation Authority, 2018a). However, they are debatable in their defensibility and are subjective dependent upon political, societal, economic and industrial conditions (Ale et al., 2015).

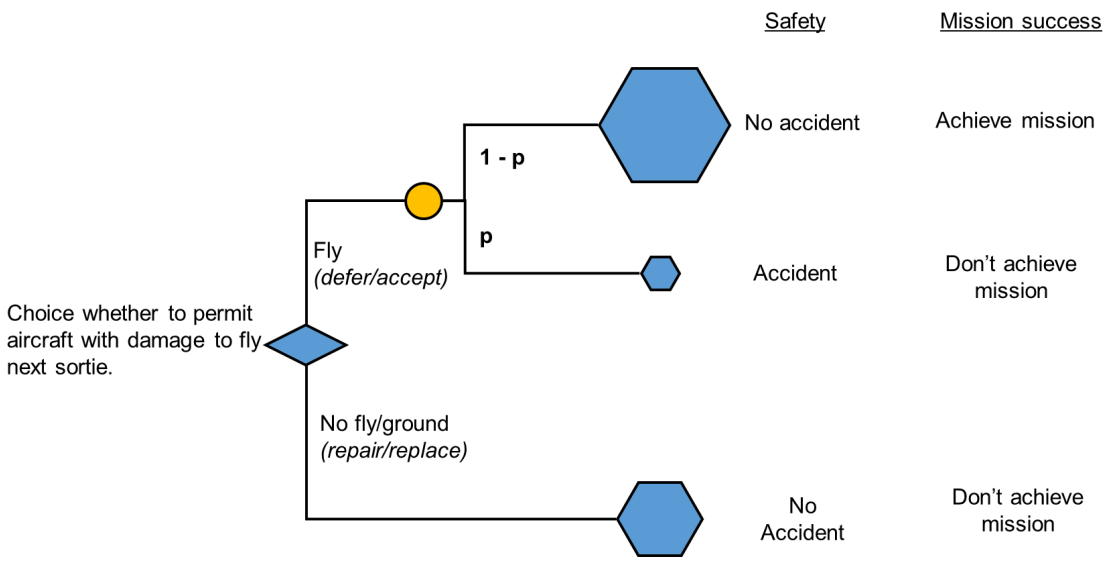


Figure 1.1: The simplified aircraft maintenance decision problem. Decision tree representation adapted from Ludlow (2020).

indicates the probability that the structural fault will lead to a loss of airworthiness on the next sortie. Probabilistic, quantified measures for risk are useful for comparing risky prospects to allocate resource (Cox, 2008), encouraging dialogue between risk stakeholders (Goerlandt et al., 2017), and for justifying the rational basis of a subsequent decision (Zio, 2018). They are easy to manipulate with risk calculus through methods such as fault and event tree analyses to calculate the overall risk to life, and allow non-expert aircraft engineering maintenance practitioners to make improved decisions in both design and operations (Bier and Cox, 2007). While probabilistic insight is easier to establish for aircraft systems that have reliable and repetitive historical performance data, such as for mechanical (hydraulic, pneumatic) or electrical systems - so called 'black box' systems, this is not the case for structural systems that transmit air and ground loads through an aircraft. Historical condition data is less coherent across aircraft of the same model type and residual structural performance varies across the aircraft within the same fleet, owing to usage, environmental exposure, human handling (from either aircrew or maintenance personnel), natural material variation, or manufacturing quality differences. These variations make damage propagation rates (and hence forecasts of failure events) uncertain and difficult to predict with accuracy. Whilst assumptions upon these variations can be accounted for during design, application

to crack growth in real structures may not be as assured in its efficacy (Tavares and de Castro, 2017, p1519).

The difficulty of assessing structural risks is exacerbated in *operational* situations characterised by: time limitations for a decision (fly or ground for maintenance); complexity of structural faults; compatibility of the methods available to assess the risk with existing risk management requirements; and a human decision-maker who is susceptible to fallacies and bias (Neumann and Politser, 1992, p28-29) but is still expected to make a choice. For maintenance decision-makers deliberating structural faults, the decision problem includes subjective management considerations alongside structural performance information (if available), as depicted in Figure 1.2. This figure depicts the maintenance decision choices available[4]: repair, replace, defer corrective action or accept the fault "as is" and continue operating. These are a direct translation of the '4 T's' model: Treat, Terminate, Tolerate, Transfer, a common risk management terminology[5]. At an elementary level, the decision-maker must determine whether they believe the structural fault detected (or suspected) will jeopardise the airworthiness of the aircraft by triggering structural collapse of the airframe during any phase of subsequent flights. In addition, they must account for any contractual requirements, regulatory considerations, maintenance options available and the availability requirements for the aircraft.

In a typical *operational* scenario for an aircraft supporting overseas military operations, a structural fault or failure may be identified during pre-flight "walk-arounds". Only timely decision-making would enable that aircraft to continue to support time-critical operations, such as 'search and rescue' or 'close air support' missions. Given the remote location and under a constrained time horizon, the maintenance manager may adopt the role of risk analyst *and* decision-maker to satisfy a compelling demand for a decision solution when without access to expertise or resources to support an objective assessment. Another typical *operational* scenario may be an engineering manager at a main base who, upon

[4]Although the choices can ultimately be reduced to the simplified "fly" or "ground" options in Figure 1.1.

[5]Hiles (2012) discusses the use of the '4 T's' in operational risk management.

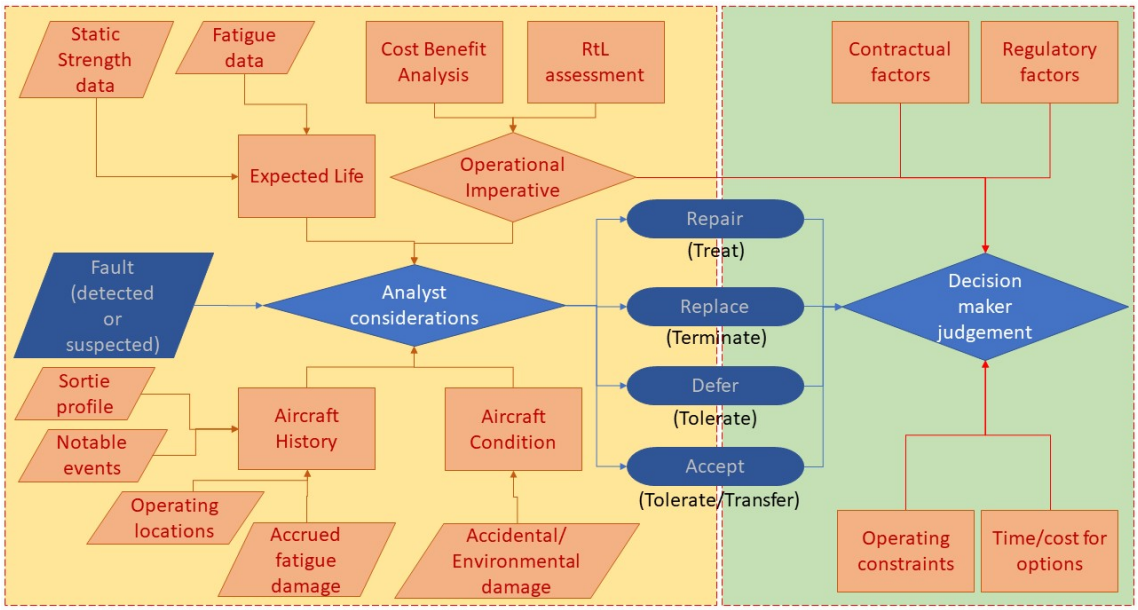


Figure 1.2: Aircraft maintenance structural integrity risk decision flow diagram for UK MOD Tornado GR4 in 2016. *RtL* is Risk to Life.

being notified of a structural fault, is responsible for making a timely judgement about the significance of a risk and whether to take intervention to safeguard the remainder of the aircraft fleet. In situations with operational constraints, it is impracticable for aircraft maintenance decision-makers who are not structural integrity experts to make analytical risk assessments about faults, and yet a decision is still required - fly, or ground. Under these circumstances, structural risks are assessed deliberatively, using subjective judgement and without support from objective analyses.

1.1 Research context

1.1.1 Risk analysis in the maintenance of UK Ministry of Defence aircraft

There have been cases reported in the public domain that highlight the strategic impact that unmitigated risks can have at commercial and political levels (Cartlidge (2015), Defence Safety Authority (2018)). In particular, the total loss of a Nimrod

Mk2 aircraft XV230 and all 14 crew owing to an explosion from an uncontained ignition risk led to a public inquiry (Haddon-Cave, 2009). The inquiry's conclusions led to a wholesale overhaul of the entire UK military aviation community's safety management framework and regulatory policy, placing increased legal accountability on individual decision-makers for any decisions that affect the safe operation of aircraft. Decision-makers, particularly in public office, must demonstrate appropriate management of risks (Military Aviation Authority (2018a), Military Aviation Authority (2018b)) throughout an aircraft's operating life since any decision leading to a negative outcome would be scrutinised to the highest level, more than likely in the public domain (Cartlidge, 2015). Decision-makers should retain evidence of their decisions in order to substantiate a selected course of action regarding risks (Public Records Act 1958, 1958). The UK military aviation community translates risks on to a matrix, with predefined intervention levels that are tied to personal accountability and seniority assignments (Military Aviation Authority, 2018a). Populating risks on the matrix necessitates a verbal or quantified measure of the consequence and likelihood for a given risk.

1.1.2 The aircraft maintenance structural risk management dilemma

Aircraft maintenance decision-makers must manage new and unfamiliar structural risks and justify intervention choices. Generating a meaningful and justifiable measure of the probability of aircraft structural failure, given an in-service fault, is too complicated a task for the situational constraints and individual decision-maker, demanding expert knowledge and model building. In order to generate a probability of failure, each influencing factor (Figure 1.2) should be quantifiably modelled, which is an unrealistic expectation for generalist engineering managers to conduct, especially under *operational* conditions. Even if structural expertise were available, a model built to objectively output structural failure risk would likely be beyond the time, information or resources available in operational situations. Until structural health monitoring systems can be relied upon to give timely, relevant, accurate and

trustworthy data concerning the health of aircraft structure, decisions regarding structural integrity will continue to require human intervention. Quantitative methods are therefore unable to be relied upon for meaningful risk measures for structural integrity risk in time-stressed situations.

From the author's experience, the practice established for managing aircraft fleet structural integrity risk on the Royal Air Force Tornado GR4 aircraft fleet[6] was for quantitative likelihood values to be stated as a subjective and notional likelihood representative of the local risk lexicon, assessed relative (and therefore approximately) to likelihood measures given for other structural risks. Consequences were judged qualitatively, and the complete risk assessment then endorsed by independent expert consensus. This translation from approximate judgements to defined quantitative risk measures illustrates the routine way in which engineering managers' and their expert consultants' judgements influence risk analysis outputs.

Decision support tools, such as risk analysis processes, must therefore be usable by decision actors justifying their choice - in the operational aircraft maintenance context outlined here, this means they must be simple and be able to be applied by a human. UK military regulatory guidance provides information about wider management of risk, about the form of risk output required and permits the use of qualitative assessment (Military Aviation Authority, 2018a). What is lacking is guidance to individual decision-makers on how they should best apply their judgement and subjective assessments to structural integrity fault likelihoods and consequences. Maintenance managers are still required to provide authoritative judgements regarding the airworthy state of aircraft, being cognisant of operational demands for aircraft availability. Defensible justification of a decision is equally as important for a third-party, as a decision in either direction (to fly or ground) may attract critical scrutiny, particularly if a decision to ground was regarded as unnecessary.

Literature addressing structural integrity risk specifically targets the objective

[6]The Tornado GR4 fast jet two-seater bomber aircraft was retired from Service from the United Kingdom Royal Air Force in April 2019.

calculation of failure probability, which necessitates time, resources and information that is unlikely to be available in operational conditions. Studies have tended to be laboratory based, where control over information sources and resources is less restrictive, and time is not a limiting factor to consider the model that should be applied to a particular structural fault. Decision-making focussed literature identifies prescriptive methods for supporting naturalistic decision-making conditions or where decision short-cuts (known as heuristics) might selectively be applied to situations. Prescriptive approaches have generally considered time-critical emergency decisions, and heuristic approaches are seemingly limited to situations where the validity of particular cues is known *a priori*. This is not the case being faced in an engineering maintenance situation. New structural faults are generally unique and information is scarce, but any decision made regarding the structural risk is made deliberatively, just without the support of formal analysis. Maintenance decision-makers are boundedly rational in making their intervention choices, meaning their rationality is limited by the environment and their cognitive limitations (Simon, 1955). However, there is no consistency in judgement-based approaches and no assurances can be made that such decisions are being made in a procedurally rational fashion (Simon, 1979).

Improvement is needed to uphold accountability, promote transparency and enable better appraised management decision-making about aircraft maintenance structural risks. The author's experience of decision-maker practice in respect of real-world structural integrity risks was the motivation for the research project. The initial intent was to identify the generality of subjective judgement use in structural integrity risk analysis and identify a means to systematically improve the decision output.

1.2 Predicting aircraft structural failure

Material failure is caused by one of, or a combination of several mechanisms, summarised in Table 1.1 (Tavares and de Castro, 2017). Scatter shown in material strength test results, caused by variability in loading, environmental

	Percentage of failures	
	Engineering components	Aircraft components
Corrosion	29	16
Fatigue	25	55
Brittle fracture	16	-
Overload	11	14
High temperature corrosion	7	2
Stress corrosion cracking/corrosion fatigue/HE	6	7
Creep	3	-
Wear/abrasion/erosion	3	6

Table 1.1: Frequency of structural failure mechanisms in engineering and aircraft components, extracted from Tavares and de Castro (2017).

factors, manufacturing processes, material properties and the analyses themselves (Tavares and de Castro, 2017), makes the ultimate failure point prediction difficult. Laboratory experiments have confirmed that the inhomogeneity of material can create large scatter when considering fatigue crack growth, and the variability can be different between manufacturers (Wu and Ni (2007), Virkler et al. (1979)). Derived material values are able to be modelled by probabilistic distributions - the loading cycle can be described by Lognormal and Weibull distributions (Yan and Liu, 2011), and the distribution of crack size by a Weibull distribution (Wu and Ni, 2007). However, a competing view is that no single distribution could describe the distribution of crack size with loading cycles owing to inherent variability between material samples (Virkler et al., 1979). Laboratory-based deterministic modelling of fracture mechanics, where control over the loading spectrum and usage exposure can be retained, allows material variables to be identified and such structural performance observations to be made. Deterministic modelling is not currently practicable on damaged *in-use* equipment for real-time decisions about fault risks.

1.2.1 Detecting and measuring material damage

Detection of damage is the first step to being able to manage the risk from structural defects. Aircraft maintenance inspections include visual (with magnification and light source), non-destructive (such as eddy current scans, or x-ray techniques) and destructive methods (dissection and microscopic inspection of the fracture surface to identify the failure mode). Visual inspections can cover broad areas of inspectable surfaces for macroscopic sources of damage but are subject to human error. Non-destructive techniques have the resolution to detect smaller cracks than visual inspections but may require significant preparation of material surfaces, have an inherent probability of detection, and require highly trained technicians. Destructive techniques are the most comprehensive, but are also costly to carry-out and have the clear downside of requiring a sacrificial sample, which may not be desirable or practicable. Recent advances in computing and intelligent use of miniaturised sensors has enabled the structural health monitoring[7] domain to rapidly grow as a useful damage identification tool, becoming a substantial research field in its own right (Tavares and de Castro, 2017, p1517).

1.2.2 Managing material failure

Table 1.1 shows that fatigue is the main source of failure in metallic structures and is likely to remain so as structure deteriorates through normal use (Tavares and de Castro, 2017). Design philosophies have evolved in order to minimise and manage fatigue driving factors in design and construction.

– *Safe Life.* "Intended to ensure that the component or structure will not develop fatigue cracks during service" and is "able to withstand a defined fatigue life, for a given load spectrum without any inspection being required" (Tavares and de Castro, 2017), "demonstrated by extensive structural ground testing including a full scale fatigue test of all major structural items" (Dilger et al., 2009). It is an "effective way to mitigate risk without necessarily quantifying just what [is] the risk", effectively making the risk negligible. It

[7] Structural health monitoring will be discussed in greater detail in Chapter 2.

was born out of the Cold War era when obsolescence (driven by technology and capability brinkmanship developments), not fatigue limits, controlled aircraft retirement timescales (Hurtado and Hoffman, 2006, p630).

– *Fail-safe.* Aims at achieving a satisfactory life with no significant damage. The structure is also designed to be inspectable in service and able to sustain significant damage before safety is compromised (Tavares and de Castro, 2017).

– *Damage Tolerance.* The most widely used philosophy in aircraft at present (Tavares and de Castro, 2017), assumes the component can continue to be safely operated in the presence of damage, which may grow in service up to a limiting value. The emphasis is upon detection and the structural design assumes that cracks will be detectable before ultimate failure, which is determined by the toughness of the material, plastic limit load or another criteria. This necessitates periodic inspections and the quantification of crack growths given applied design loads (Tavares and de Castro, 2017).

– *Limit of Validity.* This emerging philosophy concerns a specific damage phenomena called widespread fatigue damage[8] and the Federal Aviation Authority (FAA) have stipulated now that design approval holder's must be able to stipulate the period of time (flight cycles or hours) during which widespread fatigue damage will not occur, else the FAA will apply a conservative judgement (Tavares and de Castro (2017) and Federal Aviation Authority (2011)).

The recently retired Royal Air Force Panavia Tornado GR4, an ageing (40 year old) fast jet aircraft was designed to a safe life philosophy.[9] The relatively

[8] "Multiple site damage and multiple element damage cracks are typically too small initially to be reliably detected with normal inspections methods. Without intervention, these cracks will grow, and eventually compromise the structural integrity of the airplane, in a condition known as widespread fatigue damage"(Federal Aviation Authority, 2011).

[9] The Tornado GR4 was initially designed through a consortium with other German and Italian partners during the Cold War, and had it's maiden flight in 1974 and was designed as a multi-role combat aircraft. In the RAF two variants were produced - an Air Defence Variant and an Interdiction Strike variant, with profoundly different functions and flying profiles (Royal Air Force, 2019a).

modern fast jet aircraft Eurofighter Typhoon[10] is also largely safe life, selected as it met the customer requirements closest, and it is simple to use the universally applicable Miners' cumulative damage growth rule in line with British Defence design standard 00-970 (UK Ministry of Defence (2007), Dilger et al. (2009)). In comparison, United States Air Force (USAF) standards (United States Air Force, 2016) require damage tolerance for all "safety-of-flight" structure, with the use of safe life concepts being applied only on a limited basis where damage tolerance is not practicable.

Irrespective of the design criteria established for an aircraft's structure, manufacturers and their customers make assumptions on how equipment will be operated in-service during design phases, which may not remain accurate. Dilger et al. (2009) state that in-service environmental effects on fatigue are negligible owing to the applied surface protection remaining intact through-life, that their full-scale fatigue test[11] would remain representative of the mission profiles being flown throughout life, and that the aircraft would not be required beyond its designed in-service life. These conditions were all broken by the UK Royal Air Force Tornado fleet, as the operators sought to maximise value for money from the equipment during their 40 year in-service operating life. When design assumptions are broken, the opportunity for unforeseen risks to develop grows. Dilger et al. (2009) reference a 1970s study verifying the accuracy of spectra used for safe life philosophies, but provide no verification that this study remains applicable to modern aircraft. Another example is the total loss of a UK helicopter conducting military training, which was found to be a consequence of metal fatigue failure resulting from errors in assumption about the accrual of fatigue when landing on sloping terrain (Defence Safety Authority, 2018). Thus, despite designing aircraft to minimise the chances of structural damage, changes to in-use requirements will create the conditions for material damage to occur during use if they are not

[10] The Eurofighter Typhoon FGR4 was designed as a consortium with German, Italian and Spanish partners, and had it's maiden demonstrator flight in 1994, although the first production aircraft flew 2002. It was designed as a fully multi-role aircraft (Royal Air Force, 2019b).

[11] Full scale fatigue models are tested with assumed worst case loading spectra for a number of simulated flight cycles that are pre-designated factor higher than expected life, in order to identify where damage will occur first.

accounted for. Some management mitigation is offered by the recent update to the USAF standards, which require revisit to certification procedures if there are changes in operational life (United States Air Force, 2016, p2). The UK military aviation regulation necessitates a similar practice through routine reviews of the aircraft “statement of operating intent and usage” (Military Aviation Authority, 2020d, p7).

Aside from the design philosophy, material usage has also changed with modern high performance aircraft employing composite materials more ambitiously and in more critical load bearing situations. For example the Airbus A400M has a fully composite centre wing box structure (Sloan, 2012) and a majority of the Eurofighter is of carbon fibre composite construction in order to maximise the directional strength-to-weight benefits that composites can offer (Dilger et al., 2009). However, anisitropy and higher material performance scatter found in composites compared to metallic structures suggests that current structural integrity methodologies may not be appropriate for composite materials on account of a lack of accumulated experience in fracture and propagation of additive materials (Tavares and de Castro, 2017). This challenges existing assumptions, like by Dilger et al. (2009) who observed that the Eurofighter full scale fatigue test model composite damage was minor and remained stable for the duration of testing. Real-world examples raise questions regarding the gap between identification of material damage on *in-use* equipment (composite and metallic), and the management methods necessary to make meaningful judgements on the reliability and serviceability of structure, especially when a manager will be requiring answers to questions such as (Frangopol, 2011):

– Has the load capacity of the structure changed?

– What is the probability of failure of individual members and the whole structure?

– What preventative maintenance is required?

1.3 Research aim and objectives

Forming and communicating unsupported judgement-based decisions is not presently governed by any of the regulatory bodies in the aviation domain reviewed. A decision support framework for situations where analytical solutions are unavailable would deliver broad benefits. Regulators could see a decision audit trail, organisations could scrutinise available information or resources at the decision point to improve processes, line management would be able to review the detail of the decision basis and assumptions, and individual decision-makers could gain feedback on their decision and understand their own biases. The coherent linking of judgement-based operational decisions can only be done if the decision basis is transparent and the decision-maker documents the decision as a requisite representation given the information available at the time.

Two primary research fields have been identified to address the context of aircraft structural integrity damage assessment in an operational situation. With the risk and decision analysis domains already heavily populated with substantiated works, isolating an appropriate research home to support improving aircraft structural integrity decision-making in operational situations requires exploration of both areas. The primary research aim is therefore:

To support engineering decision-makers deliberating aircraft maintenance structural risks under constrained decision-making conditions.

The supporting objectives of the research project should address the following two basic requirements, which are summarised from this chapter:

1. Gain a detailed understanding of the beliefs and practices of real-world practitioner decision-makers deliberating structural risks in operationally constrained conditions.

2. Identify a method that supports decision-makers in forming a timely, effective, justifiable and robust judgement on unexpected aircraft structural damage in operational situations.

1.4 Overview of research design and methodology

The research aim concerns the human deliberating under time pressure over whether aircraft structural damage is expected to propagate to failure or not, and consequently if the damaged aircraft should be considered as safe to fly. No two human decision-makers will consider a situation in the same way given individual knowledge and experience bases. To research and design a support tool for unsupported judgement-based decision-making necessitates an interpretivist epistemology and requires the researcher to understand social actors within their own cultural setting (Wilson, 2010, p11). From an axiological perspective, this approach is suitable given the author's experience of aircraft maintenance decision-making. Thus the research must accommodate the complexity associated with decision actors and their perspectives on structural faults. Otherwise, adopting a positivist paradigm might encourage a false impression of precise and objective research data that identifies relationships and makes conclusions that are unhelpful and not generalisable. This would be complicit with a 'work-as-imagined' perspective of safety, rather than 'work-as-done' (Rae et al., 2020, p7). On this basis, a pragmatist ontology is best suited to studying how to support human decision-makers using their unsupported judgement about structural integrity aircraft risks.

To verify research outcomes, improvement must be measured. Measuring improvements in risk assessment accuracy as a consequence of decision-maker choices is not useful, since true risk values would be necessary to demonstrate accuracy. Establishing the true risk for a specific fault within complex aircraft structural systems is not possible without representative and full-scale aircraft system frequencies of failure. It would also be beyond the ethical limits of the research and would be prohibitively expensive. Improvement is therefore targeted by using peers to compare decisions made using an intervention approach against existing representative decision-making practice. If the intervention establishes a better and more transparent decision justification, and the decision actors feel positive using it, then in the subjective space of decision-making, it will be claimed

that improvement has been realised.

An abductive strategy is adopted to first verify how real-world decision-makers approach structural integrity risks in real-world operational situations and, subsequently, inform the development and test of a novel framework to improve decision-making support under such circumstances.

Research management Project management was supported by a project plan, data management plan, a risk assessment and ethical reviews related to data collection. While the latter will be covered by sections discussing Phases 2 and 4 of this research (phases defined below), the other items were essential to ensure that legal mandates and self-supportive practical aspects of the research project as a deliverable intent were achieved. A snapshot of the project Gannt chart is included in Appendix A, which indicates the delays experienced to the original planned timescale, a result of the requirement for external ethical review (refer to Section 3.1.2), which delayed initial data collection and had a subsequent knock-on impact.

1.4.1 Methodology

The study was structured into four phases:

- *Phase 1.* Literature review and identification of relevant academic research gaps.
- *Phase 2.* Initial data collection from structural integrity risk decision-makers and results analysis. Inductive identification of patterns in practitioner behaviour when handling structural integrity risks under operational constraints.
- *Phase 3.* Novel framework development to support decision-makers under operational constraints, informed by the findings from Phase 2.
- *Phase 4.* Novel framework testing, data collection and results analysis. Deductively test if a novel framework provides improved decision support

and audit evidence of subjective judgements made regarding structural integrity risks.

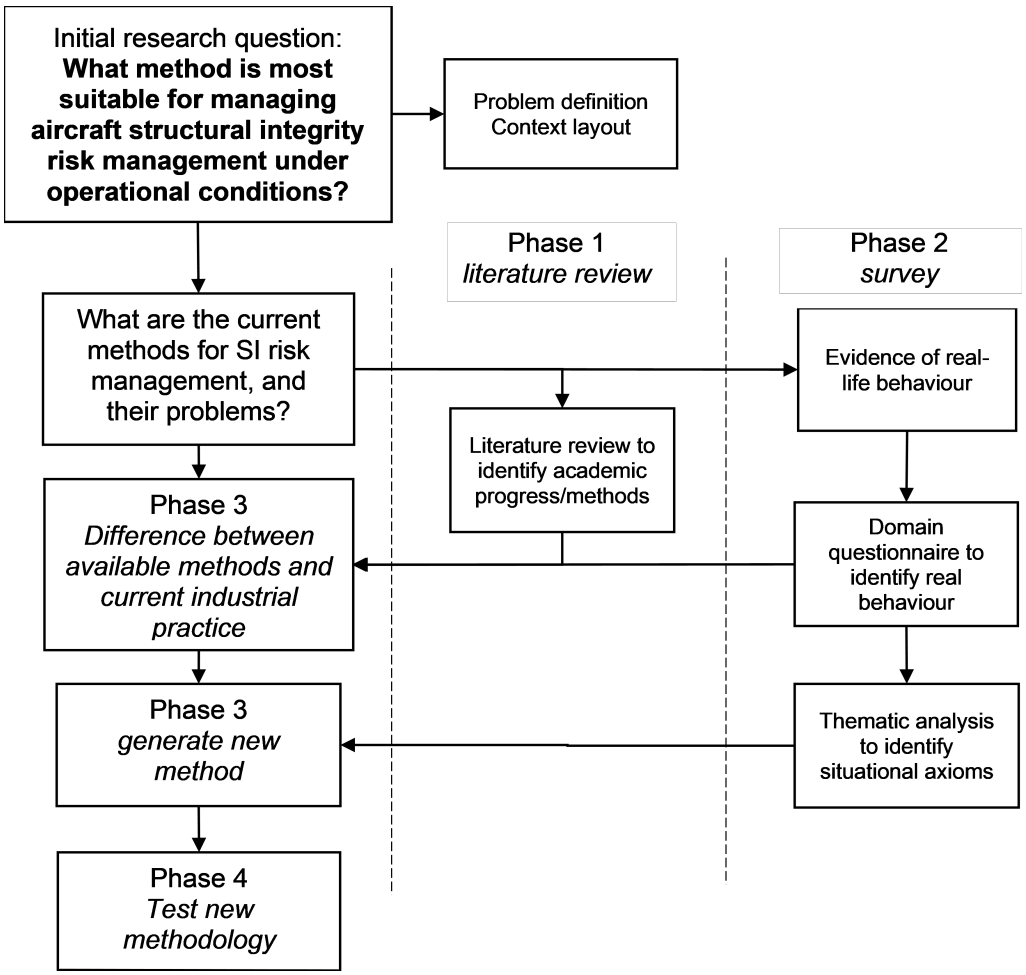


Figure 1.3: Process flow of the research project, indicating the Phase allocations.

Figure 1.3 represents how the research project was broken into tasks and how the individual phases interact.

1.5 Thesis layout

The phases in Figure 1.3 show the chronology of the research project, and consequently the layout of the thesis follows suit. The layout steps the reader through the linear project process whereby the research question was narrowed following the literature review in Phase 1 and the survey and evidence of real-world behaviour in Phase 2. A compatible solution to the refined research question was

then derived in Phase 3 and tested with representative participants in Phase 4. The data gathering in Phases 2 and 4 both necessitated individual methodological justification, owing to the abductive approach chosen for the project. Consequently, these phases contain their own independent methodology discussion.

The chapters of the thesis are described in the following. A review of literature relevant to aircraft structural integrity risk decision-making, undertaken in Chapter 2, identified that there were no prescriptive risk decision approaches to support practitioners in the type of constrained conditions considered here. A survey of engineering risk practitioners and case study analysis, outlined in Chapter 3, highlighted that practitioners rely heavily upon unsupported judgement for deliberating and managing structural integrity risks, the results of which are discussed in Chapter 4. These results combined with the literature review informed the development of a hybrid novel framework for supporting maintenance decision-makers compelled to deliberate an aircraft structural fault under constraints of time, information and resource. The framework is based upon a heuristic decision tree for alternative selection and a justification approach that uses the Toulmin argument schema. The heuristic tree presents three "gates" that rapidly terminate to an action that either grounds the aircraft (repair, seek further advice) or permits the continuation of flight. The Toulmin schema provides a structure for decision actors to arrange available information, knowledge and perceptions in order to justify the decision choice, enabling scrutiny of the information acted upon and a clear audit trail. Chapter 5 describes the novel framework development and testing strategy. The results, presented in Chapter 6, indicate that the novel framework provides improved decision transparency and a better justification when compared with existing decision-maker approaches. Chapter 7 presents conclusions of the research project and some final remarks.

1.6 Chapter summary

This context chapter has outlined that quantifying aircraft structural integrity risks in an operational situation is difficult and time consuming. A changed emphasis on risk in UK Ministry of Defence aviation regulation requires that all issues that have the potential to reduce the safe operating level of military aircraft must be risk assessed relative to pre-determined safe threshold values. This research project focusses on decision-making situations where operational circumstances compel deliberations about aircraft structural faults without the opportunity to optimise decisions with analytical methods.

The project aims to investigate naturalistic engineering decision-making situations through an abductive research strategy. Overarching objectives are to improve understanding about industrial practices to managing structural integrity risk, particularly under operational constraints, and develop a methodology that can better support decision-makers justify judgement-based deliberations about aircraft structural faults and failures.

Chapter 2

Literature review

"*. . . actions require beliefs, but beliefs do not necessitate action*" (Lindley, 2014)

The key points from Chapter 1 enable a set of pseudo-requirements to be established for the context framed by the UK Ministry of Defence aviation sector, which establishes a clear focus for the literature review. These are the *contextual requirements*:

1. The decision context concerns *aircraft engineering maintenance* situations, where the decision presides over a *situation, regarding aircraft structure, that has the potential to cause harm or loss to stakeholders*. This may be following the identification of damage (e.g. a crack in a frame or corrosion to material surface), or on receipt of information that damage may exist on the aircraft (but cannot be inspected or examined).

2. Decisions being made regard equipment that has entered service and is *in-use*.

3. The decision environment is *operational*, where there may be insufficient time, resource or information to complete an analytical model of the situation.

4. The output of the decision concerns selection of the most appropriate course of action, given the known time horizon, available resources and state of knowledge that constrain the decision in the decision environment.

5. Decision bases are required to be *auditable* and should fully represent the situation at the point of the decision from the perspective of the decision-maker.

6. Decision-makers are generalist aircraft engineering managers who must decide what action to take; tools for decision support must be usable by this cohort.

This chapter is split into two major sections in order to review the two domains of risk and decision analysis with relevance to the industrial need identified in Chapter 1. An initial discussion regarding safety science frames the discussion of risk and decision analysis from a perspective of safety, as opposed to financial or business perspectives.

2.1 Aviation safety

Consideration of the interactions that may result in harm to humans necessarily requires comment about safety science. This field has sought to understand and prevent accidents, proposing theories to guide industrial practice and the development of preventative measures (Dekker, 2019, p1). It is a multi-disciplinary field spanning psychology, social science, physical sciences, engineering and population health. The literature review does not include an explicit review of the "safety assessment" domain (although it will be referred to), since this domain contains holistic aspects such as organisational and process mechanisms to ensure and assure safe standards of operation in an organisation, which are outside the context of this study. Although, it is acknowledged that safety colours the problem space being explored.

The direction of this research project complements advances being made in resilience engineering, a modern safety paradigm that aims to help people cope with complexity under pressure to achieve success (Dekker, 2019, p402). Aircraft structure is a complex system with numerous uncertain human, technical and environmental interactions that influence performance. Although total aircraft

system failures are, thankfully, rare occurrences as a result of upstream advances in design practices, the shortfall in advice about how practitioners should apply their subjective judgement during structural risk deliberations compels proactive improvement. Should an accident occur as a result of a misjudgement about structural risk, the moral regret would be inescapable.

Safety science thinking has elevated human factors considerations into real-world practice, and training systems and regulatory guidance material can be found highlighting how technology and humans can interact in an accident condition (for example Civil Aviation Authority (2003) and Civil Aviation Authority (2002)). Extant industrial guidance provides awareness of human judgement frailties and is configured for accident avoidance and addressing causal factors, such as managing fatigue to prevent poor decision-making (Civil Aviation Authority, 2003, Chapter 4, p3-4). However, in operational situations, control over such conditions may be restricted and there are no instructions on what should be done when a decision is still necessary. These are the conditions which this research project is supporting, with specific regard to structural integrity risk decision-making.

2.1.1 Military aviation management strategy for structural integrity

Mitigating the difficulties in managing structural integrity risk and uncertainty using traditional methods, the UK Ministry of Defence follow a comprehensive whole-life approach to managing in-service aircraft. The ‘ESVRE’ framework is an assurance framework that covers initial design and through-life structural considerations and is mandated by the UK military flying regulatory body, the Military Aviation Authority (MAA) (Military Aviation Authority, 2020d). The framework requires that all aircraft covered by the MAA’s regulatory scope are able to demonstrate auditable processes for ‘establishing’, ‘sustaining’, ‘validating’, ‘recovering’ and ‘exploiting’ structural integrity. The framework represents proactive management activity in order to mitigate against factors in structural integrity failure, such as environmental

and human factors, while ensuring that processes are in place to prevent, detect and recover damaged structure before aircraft integrity failure. Similarly, the USAF operate a 5 stage process in their Aircraft Structural Integrity Program (ASIP) (United States Air Force, 2016), which covers whole-life considerations and is based on sustaining accurate analytical structural performance data updated with in-service changes. This Standard assumes a complete data set and makes no consideration for human decision-making in circumstances where the data set is incomplete.

Current military maintenance policy does not appear to provide guidance to personnel on how to apply their judgement to risk assessment. In the UK military aviation domain, maintenance managers are permitted to release an aircraft for flight provided their assessment about structural integrity risk is "considered justifiable and safe"(Military Aviation Authority, 2020b). In open source USAF maintenance policy[1], only broad references could be identified to maintenance decisions about faults which regarded "deferred discrepancies" (US Department of the Air Force, 2020). General risk management policies in both domains provides scope for qualitative risk assessments, acknowledging that some situations necessitate individualistic perspectives (Military Aviation Authority (2018a), US Department of the Air Force (2019)), but no prescriptive guidance is found in any military aviation maintenance policy reviewed.

2.2 Risk domain review

Discussion is focused on recent literature relevant to operational decision-making of structural integrity risk and includes both domain-specific and broader literature. The regulatory environment described for the context of this study requires a measure of structural integrity risk to be established, in order to provide assurance of aircraft airworthiness levels against a pre-determined acceptable threshold. Methods that use frequentist analysis of data are largely discounted for structural

[1] It is acknowledged that some internal procedures may not be accessible owing to public access limitations.

integrity risks identified in-service since they depend upon data drawn from repetitive scenarios. Subjective and Bayesian-based approaches are advocated as suitable for use in low and poor information settings, but their usability in operational situations is unverified.

2.2.1 Paradigms of risk

Risk analysis is established as a practice and has grown as a substantive research field from the mid 1970s and early 1980s when the first editorial of the journal *Risk Analysis* was published. Coherent progress has been hindered by stove-piped research regarding domain-specific issues, which has created specialized competencies and inconsistencies across interdisciplinary boundaries such as medicine, oil and gas, process industries, and aerospace (Hansson and Aven, 2014). There are numerous interpretations of 'risk' lexicon ((Aven and Zio, 2014) and (Aven, 2016b)) and the credibility of risk analysis as a scientific field has required justification (Hansson and Aven, 2014). The plethora of multi-disciplinary risk research that has proliferated into industry to represent real-world systems probabilistically has undoubtedly meant that risk analysis has improved safety by increasing formal organisational discussion on events that could create harm or loss. A counter-productive consequence is that some inherent weaknesses have been normalised in real-world risk management applications, as has been seen by the use of risk matrices (Cox, 2008). The probabilistic representation for risk is most common, so a brief introduction on probability is appropriate for understanding later discussions about literature. The starting point for this discussion is the unifying set of axioms first defined in 1933 by Kolmogorov (2018, p2), which are described for events A and B, within a set of events S:

- The probability of event A, $\mathbf{Pr}(A)$ is a non-negative, real number, i.e. $\mathbf{Pr}(A) \geq 0$.
- In a set S of all possible events, the sum of all event probabilities (e.g. $\mathbf{Pr}(A), \mathbf{Pr}(B)$etc.) within the set is 1, i.e. $\mathbf{Pr}(S) = 1$.

- If events A and B are mutually exclusive, then the probability of them both occuring, i.e. their *union*, is $\mathbf{Pr}(A \cup B) = \mathbf{Pr}(A) + \mathbf{Pr}(B)$.

Classical Probabilistic Representation The simplest explanation of probability is the classical definition, which concerns observable measures of equal likelihood, such as those presented by a fair coin, a fair dice, or a fair deck of cards. The probability of a 'heads' from a fair coin is $\frac{1}{2}$, or 50%, derived generally as follows:

$$\begin{aligned}\mathbf{Pr}(\text{Event of interest}) &= \frac{\text{Number of times event of interest occurs}}{\text{Number of times relevant events occur}} \\ &= \frac{f}{N}\end{aligned}$$

The classical definition is not helpful for assessing the risk of an aircraft structural fault. Although useful for understanding probability in relative terms, the definition is absolute and does not allow for the inclusion of occurrence data or belief. For this, alternative representations of probability are required.

Probabilistic Representation of Belief A probability can also describe a numerical representation of an individual's degree of belief (or uncertainty) in a variable, which can then be mathematically manipulated, by logic and reason, to assist in calculating a prediction for a state. But it does not actually exist as an observable parameter (Lindley, 2014). One probabilistic convention is to consider an urn containing two sets of coloured balls; for ease the combined total number of balls is 100. A trustworthy source states that there are 50 red and 50 blue balls in the urn; the probability of pulling a Blue Ball from the urn is the fraction of Blue Balls to the total number.

Using these principles, belief for any given situation can be quantified comparatively by using balls in an urn as a hypothetical baseline - "is it more likely to Rain Tomorrow than for me to remove a Blue Ball from an urn when $\mathbf{Pr}(\text{Blue Ball}) = 0.5$?" If the answer is no, then further hypothetical questions are asked - "is it more likely to Rain Tomorrow than for me to remove a Blue Ball from the urn when $\mathbf{Pr}(\text{Blue Ball}) = 0.8\ldots$ or $\ldots 0.7\ldots$ or $\ldots 0.75$?" The probability

of removing a Blue Ball is adjusted until the point at which *you* are indifferent between removing a Blue Ball and Rain Tomorrow. In this way, *your* belief in an uncertain event can be represented as a probability, or if preferred by using an odds format[2].

This type of probabilistic statement is conditional on individualistic subjectivity, usually indicated by a term that represents our current state of knowledge - $\&$[3]. This notation is useful for advising that a probabilistic risk analysis is conditional on the accuracy of individual knowledge, but it is rarely seen.

2.2.1.1 Risk and engineering

Techniques for predicting outcomes in engineering systems fall into three paradigms covering probabilistic and non-probabilistic methods (Hansson and Aven, 2014): analytical paradigm (non-probabilistic), frequentist paradigm (probabilistic), and subjective probabilistic paradigm (probabilistic).

Analytical methods Analytical methods, or physics-based approaches (Animah and Shafiee, 2018), rely upon deterministic formulations from elementary measures of the structure, material and situation being modelled. Design data can be established for structural components by controlled experiments covering expected loading, environmental assumptions, computer-based modelling such as Finite Element Methods, and full-scale fatigue test models which validate design assumptions with expected loads. Safety factors[4], quality assurance, inspection and maintenance (Theofanous, 2003) account for in-service differences in loading and inherent material parameter variability (Wu and Ni, 2007).

[2]Representing belief by using odds: "is it three times as likely to remove a Blue Ball from the urn, than it is to Rain Tomorrow?" This representation could then be described as a ratio 3 to 1. One unit of certainty for Rain Tomorrow against 3 units of certainty for removing a Blue Ball corresponds to $\mathbf{Pr}(\text{Blue Ball}) = 0.75$ and $\mathbf{Pr}(RainTomorrow) = 0.25$.

[3]The probabilistic statements would then be $\mathbf{Pr}(\text{Blue Ball}|\&) = 0.75$ and $\mathbf{Pr}(RainTomorrow|\&) = 0.25$, which reads as: the probability of Blue Ball *given* (indicated by the symbol for conditional probability $|$) the current state of knowledge $\&$ is equal to 0.75; and the probability of $RainTomorrow$ *given* the current state of knowledge $\&$ is equal to 0.25.

[4]The UK Defence Standard for aircraft design necessitates the use of safety factors when calculating fatigue lives (UK Ministry of Defence, 2007).

Once equipment is in-use, analytical methods cannot be proven accurate for risk assessing without knowing the loading history and material state parameters at precisely that time. Controlled trials with many repetitions of the situation being faced could provide statistical evidence of unknown variables (Rae et al., 2014), but this is not practicable and as Weinberg outlined (quoted in Rae et al. (2014)) "one of the most powerful methods of science - experimental observations - is inapplicable to the estimate of overall risk". Applying conservative assumptions is a tempting alternative (Aven, 2016a), or carrying-out preventative action prior to firm scientific evidence (Spiegelhalter and Riesch, 2011) to treat situations with insufficient information, but these approaches cloud the precision of predictions. By measuring and modelling physical behaviour of structures, structural health monitoring (SHM)[5] will become crucial to the provision of accurate through-life monitoring. As in-service SHM continues to permeate into complex modern system designs, analytical methods may become useful in risks analysis in the future.

Frequentist Paradigm Frequentist probabilities are built upon the assumption that repetitive observations of independent and identically distributed variables will produce a long-run relative frequency, which can be interpreted as probabilities and used for predictive inference. Processes or systems that have unambiguous metrics representative of system life can apply specific criteria such as hours of usage, or cycles. Historical usage data can then allow the prediction of future performance, provided there is sufficient identical and independent data, for example:

$$\mathbf{Pr}(\text{Component Failure}) = \frac{\text{No. Component Failures}}{\text{Total no. operating hours across all similar components}}$$

The requirement to hold sufficient statistically similar data makes the frequentist approach unsuitable for predicting future structural integrity occurrences owing to uniqueness as a result of material aleatory uncertainties. For example, a structural failure on training aircraft may not be analogous to front-line aircraft flying different

[5]Structural health monitoring will be discussed in greater depth in Section 2.2.5.1, but it broadly has the objective of damage detection and assessment (Thöns, 2018, p79).

sortie profiles. Fictitious populations modelled with assumptions about past and future use would be necessary if one was to define a statistically similar data set for structure, but this does not provide a meaningful representation of the situation (Aven and Zio, 2011). Furthermore, in-use equipment usage may differ from that originally intended at initial design, which may challenge the efficacy of design information used as *a priori* data. These variations inevitably mean that the number of representative and repeatable structural fault arisings are insufficient to provide statistical confidence in frequentist predictions.

Subjective Probability Paradigm Subjective methods (also known as the Bayesian approach (Aven and Zio, 2011)) offer alternative probabilistic approaches by being reliant upon human-prescribed measures of belief, usually elicited from experts. Subjective probabilities are wholly influenced by belief, human behaviour and individual risk propensity, which introduces psychological and environmental contributory factors. Most modern accounts of subjective probability use the theory of Subjective Expected Utility as a base (Section 2.3.1.2), which assumes that humans are rational instruments. Indeed, some understand that "one's degree of belief must be rational" if it is to be considered as a probability (Peterson, 2009). Subjective methods can be conducted individually or by a group of experts, to elicit a range of possible values for an uncertain parameter or possibility of an event (Oberkampf and Helton, 2002). While some approaches are more popular than others, no single obviously best elicitation method exists across all application possibilities.

Bayesian probabilistic representation Bayesian representations capture individual belief given known information, adopting a parameter that encodes a personal perspective upon the situation being described by a probability. The use of Bayesian methods uses the *conditional* notation ($|$), for example $\mathbf{Pr}(B|A)$ is read as the probability that event B occurs given that event A has already occurred.

Bayes' formula enables the calculation of a conditional probability using a 'likelihood' parameter and a 'prior distribution' parameter that are specific to a

particular *situation*. For effect Y, and cause X_i, where i is 1 of n possible causes, the Bayes' formula is as follows (Lambert, 2018, p12-13):

$$\mathbf{Pr}(X_i|Y) = \frac{\mathbf{Pr}(Y|X_i)\mathbf{Pr}(X_i)}{\mathbf{Pr}(Y)} \tag{2.1}$$

The dependent parameter in Equation 2.1, $\mathbf{Pr}(X_i|Y)$ is known as the *posterior* probability. The expression for 'likelihood'[6] ($\mathbf{Pr}(Y|X_i)$) describes the information known already about the *situation* as a probability. The expression for the 'prior' ($\mathbf{Pr}(X_i)$) describes the belief in X_i before Y is observed. The denominator ($\mathbf{Pr}(Y)$) is the normalising parameter, representing sum of all Y values given X_i and $\bar{X}_i$[7]. Bayes' probabilistic representation is useful precisely because it allows inferences to be drawn from the information currently available, meaning it can be updated as new information becomes apparent. For more in-depth explanations, refer to Jordaan (2005, p48) or Clemen and Reilly (2001, p255). However, its suitability for operational situations that are time constrained and information poor may render its usability unattractive to generalist engineering managers.

2.2.1.2 Uncertainty and probability

Probability as a concept is a characterisation of uncertainty. Probabilistic Risk Analysis (PRA)[8] approaches comprehensively and systematically evaluate uncertainties, which can result in overconfidence that every situation can be solved by an applicable 'theory' or model (Theofanous, 2003). In order to improve risk analysis trustworthiness and credibility, it is essential to understand the nature of uncertainty (Pasman and Rogers, 2018), which can be anywhere between total ignorance about parameters, to objective chances that are well documented for a particular parameter (Yates and Stone, 1992, p14). There may be uncertainty about which artefacts would be exposed to a hazard, about the degree to which

[6] In the Bayes' formula context, the likelihood parameter gives the probability of generating the Y if the parameters in the model were equal to X_i (Lambert, 2018, p23).

[7] $\bar{X}_i$ means "not X_i", representing all values that are not in the set of values associated by X_i, for example X_j, or X_k.

[8] Probabilistic Risk Analysis is also known as Quantitative Risk Analysis (QRA) and Probabilistic Safety Analysis (Zio, 2018); they are types Quantitative-based approaches to Risk Analysis.

those artefacts would be impacted, about the likelihood of a hazard occurring, or about the effectiveness of planned mitigations. Sources of uncertainty have been categorised as follows:

- *Epistemic uncertainty.* Concerns a lack of knowledge about a phenomenon but more information would reduce this uncertainty, for instance carrying-out non-destructive testing on a metal bracket in order to improve understanding of the dimensions of a crack.
- *Aleatory uncertainty.* Concerns random variation in a population that can be ascribed to a real distribution, which cannot be reduced owing to the intrinsic variation in parameters (Theofanous, 2003). For example material parameters can be altered away from their once known state by manufacturing treatment, impurities in the material itself, by environmental exposure, or by human interaction. Aleatory is also referred to as irreducible uncertainty, stochastic uncertainty, or chance.
- *Ambiguity.* A third source that is less widely adopted in risk analysis contexts but is more applicable to decision analysis is ambiguity, which Brugnach et al. (2010) stated as being a behavioural aspect and therefore allows uncertainty representation to be inclusive of the human element. Deterministic representations of uncertainty are less inclusive of the socio-technical environmental systems that influence decision situations, but the inclusion of the human element of uncertainty may improve dialogue between decision actors (Brugnach et al., 2010).

Other lexicons exist: Bodde et al. (2018) group by ‘inherent’(aleatory), ‘scientific’(epistemic), ‘social’(values and perceptions), and ‘legal’. There have been attempts to categorize the extent of uncertainty (Cox, 2012) and defining uncertain variables as aleatory or epistemic (Kiureghian and Ditlevsen, 2009), but there is a lack of consensus, particularly when considering the strength of knowledge characterisation in risk assessment (Aven, 2016b).

It is generally agreed that separating sources of uncertainty into aleatory and epistemic uncertainty is useful (Bier and Cox, 2007), treating each source

individually to identify where uncertainty can be reduced. Variables considered to be sources of aleatory uncertainty are then handled by considering the population to be large and similar (but not identical)[9] in order to model the variable by probability models or by using frequentist-type approaches (which assumes the model has some sort of stability) (Aven and Zio, 2014). Epistemic uncertainty has most commonly been treated using Bayesian subjective probability approaches (although assigning subjective values remains a topic of debate (Aven, 2016b)) but other methods, including qualitative methods and possibilistic measures, have been used.

Encoding individual belief in a particular (uncertain) parameter has been variously formulated using, for example, evidence theory, fuzzy set theory and interval analysis for situations that are well characterised (for example by probability distributions), or of near-total ignorance (Oberkampf and Helton, 2002). However, these representations of uncertainty are acknowledged to be difficult to understand in the context of risk assessment (Aven and Zio, 2011). They are found in applied settings (Section 2.2.5) but are criticised by many statisticians for their use of ad hoc operators. They also necessitate competent mathematical modelling, which is incompatible with requirement 6 at the start of this Chapter.

2.2.2 Reviewing the Foundations of Risk

Establishing an agreed definition for a subject matter is crucial to ensuring coherence, consistency, and clear communication. Whilst risk as a philosophy has been adopted widely across industries, an undesirable side effect is that multiple functional lexicons and definitions of risk exist across research boundaries. This weakens the academic foundations for risk assessment and management (Aven and Zio (2014), Villa et al. (2016)). Risk has been variously defined as: an expected loss; a likelihood of hazard effect; uncertainty of an outcome; the potential for a negative consequence; a combination of probability and consequence (Villa et al.,

[9] 'Similar' in this context means either 'exchangeable'/unaffected by order for a Bayesian approach, or independent and identically distributed for a traditional statistical approach (Bjerga et al., 2014).

2016); or as a triplet expression containing consequence, an uncertainty measure, and the background knowledge (Zio, 2018). Risk has also been expressed as either individual or societal risk (Pasman et al., 2017). Villa et al. (2016) provides a thorough examination of the impact of differing 'shades' of risk on individual interpretations. This lack of cohesion in definition has proliferated because of the ease of accessibility of risk-based thinking and management methods that improve perceptions of safety. Every individual will have, at some time, been confident of their predictions of future outcomes regarding risks to industrial, domestic, commercial, health, or business interests. Even deciding upon whether to take an umbrella requires a brief deliberation about the risk of getting rained upon. Interpretive differences impact real-world analysis, the subsequent decisions (Aven, 2016b) and individual perceptions of risk (Pitz, 1992, p286) - ultimately, a misinterpretation of how a risk is portrayed may result in misplaced judgement, a lack of scrutiny of analysis, or overconfidence in an analysis. The latest version of ISO 31000:2018 (The British Standards Institution, 2018a) provides a coherent definition for risk that is recent, comprehensive and flexible (Proenca et al., 2017):

"risk is the effect of uncertainty on objectives"

This definition links the influence of *uncertainty* to risk but is able to be interpreted in different ways (Aven, 2016b). Individual organisations tailor the risk parameter lexicon for their own needs, for *probability*, *severity* of adverse effects (Aven and Zio, 2011), 'risk sources' and 'potential events'[10]. The ISO 31000 definition provided for 'Likelihood' (The British Standards Institution, 2018a) further highlights the flexibility afforded to organisations in selecting risk methods and approaches that suit their requirements:

> *"Likelihood - the chance of something happening, whether defined, measured or determined objectively or subjectively, qualitatively or quantitatively, and described using general terms or mathematically (such as a probability or a frequency over a given time period)."*

[10] See the definition notes of The British Standards Institution (2018a).

Severity						
Catastrophic	5	5 Review	10 Unacceptable	15 Unacceptable	20 Unacceptable	25 Unacceptable
Hazardous	4	4 Acceptable	8 Review	12 Unacceptable	16 Unacceptable	20 Unacceptable
Major	3	3 Acceptable	6 Review	9 Review	12 Unacceptable	15 Unacceptable
Minor	2	2 Acceptable	4 Acceptable	6 Review	8 Review	10 Unacceptable
Negligible	1	1 Acceptable	2 Acceptable	3 Acceptable	4 Acceptable	5 Review
		Extremely improbable	**Improbable**	**Remote**	**Occasional**	**Frequent**
		1	2	3	4	5
		Likelihood				

Figure 2.1: Example of a Risk Matrix defined by an aviation regulatory body, extracted from Civil Aviation Authority (2015).

The latitudes afforded enable industries to develop bespoke risk management techniques tailored to their situations, business priorities, market trends, regulatory requirements, or devolved risk accountability and appetite. One such industrial example with regards to the Royal Air Force Tornado GR4 aircraft was described in Section 1.1.2. A common industrial practice is to describe a 'risk factor' in a matrix form (Figure 2.1) for ease of management using the formula (Hagmann, 2012):

Risk = *['Probability of Occurrence' or 'Likelihood']* × *['Severity' or 'Consequences']*

This format provides a concise means for communicating the output of a risk analysis (completed by qualitative or quantitative means) and comparing across multiple risks to support management allocation of resources. Despite being a simple, convenient and popular format, presenting risk information in this manner obscures important aspects of the data such as cost constraints, interactions between mitigation measures, or psychological information about the person conducting the risk categorization (Cox, 2012). Presentation within a risk matrix can also be misleading. For example for risks with negatively correlated consequences and likelihood, as in structural integrity situations (low likelihood,

high consequence), risk matrices can be “worse than useless” (Cox, 2012) and may result in inappropriate allocation of resources.

The inherent problems described with risk matrices highlight more general concerns regarding the importance of understanding the hidden elements of risk beyond likelihood and consequence, notably the uncertainty, assumptions, and state of knowledge. The individualistic nature of risk assessments means they are contingent on the choices made by the analyst (Durodié, 2017), and there is increasing desire to improve transparency in risk assessment that is consistent with the motivations of this study (Zio (2018), Lathrop and Ezell (2017), Aven (2017), Isaksen and McNaught (2019)). Transparency of individual risk characterisation is therefore important, and the trivial example below describes this metaphorically:

> *Two individuals are given the weather forecast and must carry-out a risk assessment to make their choice regarding taking an umbrella or not - one is heading to a date, the other is walking the dog. The risk of becoming wet is very important to the individual on their way to a date who needs to maintain their appearance, so their choice to take an umbrella is highly influenced by their objective of impressing their date. The dog walker may see little consequence of their becoming wet, may perceive carrying an umbrella as a hindrance and may make a different choice as a result. In both cases, value aspects within their individual risk assessments directly impacted their decision. Internalised assumptions were also likely to have bearing, such as believing the weather forecast to be accurate, which may not always be the case. The forecaster’s knowledge base may be out of date, or might be made upon an unfamiliar weather system. From outside the decision-maker’s perspective, a bystander seeing someone carrying an umbrella when it’s not raining may consider that individual to be risk averse, well prepared, poorly informed, care free, underequipped, or some other perspective.*

There is long standing awareness of individual influence and conflicts that occur

when making decisions about risk (Pitz, 1992, p304), but a lack of prescriptive guidance in practical settings infers there remains a requirement for well founded risk decision-making support tools that can be applied in restricted conditions. By understanding the values, motivations and information base of individual risk analysts, scrutiny and commentary may be more measurably considered. If the use of individual subjective judgement is widespread in the assessment and decision-making regarding a risky prospect, (as is anticipated in aircraft structural integrity maintenance decision-making) then improving risk assessment transparency will benefit risk practices and shape research focus.

2.2.2.1 Risk credibility, validity, and reliability

The risk field is challenged by a number of fundamental issues underscored by questions regarding the credibility of risk as a science, catalysed by recent public case studies where risk analyses resulted in undesirable consequences. When earthquake scientists were convicted of manslaughter in 2012[11], the assessments they provided to the Italian Major Risks Commission were found to be "unjustifiably reassuring" prior to a deadly earthquake in L'Aquila in April 2009 (Cartlidge, 2015). The UK military aviation sector is similarly answerable to public enquiry into risk decision-making (Haddon-Cave, 2009). The L'Aquila case also reveals that industry lags behind academic risk domain advances (Aven, 2016b). However, recent academic perspectives which argue for greater empirical grounding in safety research suggests this lag is creating tensions in both communities (Rae et al., 2020). Rasmussen observed two decades ago that structural changes in organisations were being outpaced by technology changes, identifying a need to systematically handle risk analyses: "it is no longer acceptable that predictive safety analysis is considered an art" (Rasmussen, 1997). Managers who became focussed on ensuring business survival were consumed by short term goals, losing sight of the changing dynamics within their organisations, leading to system safety management threats. This context (and hence his argument) remains relevant today (Dallat et al., 2019), yet these enduring challenges to the scientific

[11] The convicted scientists were subsequently acquitted in November 2014.

credibility of risk analysis (Aven, 2016b) have done little to impact their use in industry (Rae et al., 2014). These gaps and lags should be reduced to improve the resilience of the risk domain to attacks on its credibility, and improve consistency of understanding regarding risk analysis.

Risk research can be classified as either (Aven, 2016b): 1) using risk assessment and management to study and treat risk of specific activities, or 2) to perform generic risk research and development. This thesis falls into the former category but aims to support the development of wider risk domain issues by addressing the application of judgement in risk management. In order to be considered scientific, contributions to specific risk fields must provide "new insights, for example a better understanding of how to conduct a specific risk assessment method in practice" (Aven, 2016b, p2). However, a basic scientific requirement is that analysis is solid or sound, following standard protocols for scientific work, and meeting criteria for reliability and validity. In general, risk analysis as a field and the results from risk analysis methods do not adequately meet scientific requirements for validity or reliability, predominantly owing to situations of insufficient data when uncertainties were considered (Aven and Heide, 2009).

Quantitative-based risk analyses (QRA) are popular, but their outputs are difficult to validate and may mislead decision-makers (Goerlandt et al., 2017). The lack of research into the validity of quantitative risk methods is surprising (Goerlandt et al., 2017), with implications for the claim of scientific validity for risk and to improving safety. *Validity* here means either: 'conceptual validity' - the condition where an operationalisation of a concept measures what it intends to measure (as used by Aven and Heide (2009)); or 'pragmatic validity' - the condition where a method meets its intended requirements in terms of the results obtained (as used by Rae et al. (2014)). There is inadequate evidence regarding the validity of available risk analysis models, and no work addressing the effectiveness of exercises that benchmark quantitative methods. Specifically, claims of 'accuracy' and 'cost-effectiveness' are unsubstantiated owing to the difficulty of obtaining reliable results from real-world industrial evidence of quantitative risk analysis. 'Usefulness' might be claimed, owing to the promotion of discussion between

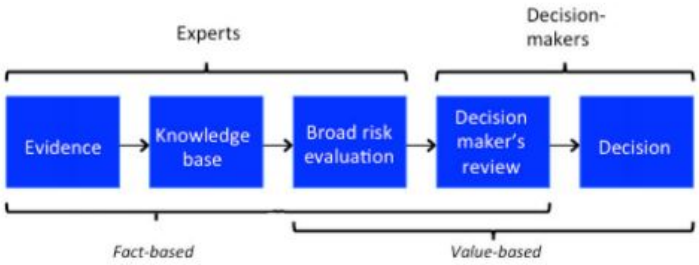


Figure 2.2: Methodology for risk analysis, separating the fact-based assessment by the risk analyst and value-based judgement by the decision-maker, extracted from Hansson and Aven (2014)

stakeholders (Goerlandt et al., 2017), but this conclusion is also debatable since there is no measure for the scrutiny that should be applied to a risk with unsubstantiated accuracy. This has consequences for circumstances that lack "absolute measures of safety"[12]. Unsafe decision conditions may prevail where the true risk threshold is unexpectedly breached. If the failing is publicly visible, this may further harm the trust in risk analysis as a whole. Bayesian-based methods[13] scored slightly better than relative frequency-based methods[14] (Goerlandt et al., 2017) against reliability and validity criteria, but this could not be verified against *true risk*[15]. Risk, it seems, cannot be validated or claimed to be reliable unless a value for true risk can be defined and measured. For rare and uncertain occurrences, like structural failures, scientific justification for the use of risk analysis is challenged.

Separating value-based decision-making from fact-based risk analysis improves the general argument for the validity of risk-informed decision-making (Hansson and Aven, 2014). Figure 2.2 shows a model for linking the individual stages, which presumes that the analysis process is reliable, valid, justifiable and

[12]An absolute measure of safety is the threshold risk level above which an event is deemed unsafe (Goerlandt et al., 2017).

[13]Bayesian-based approaches treat probability as a measure of uncertainty about future events and outcomes as seen through the eyes of the assessor given a certain background knowledge.

[14]Relative-Frequency-based approaches use a hypothetical or observed frequency of occurrence for an event.

[15]The *true* value for risk is dependent upon whether the situation is measurable using observable quantities, or using parameters of a probabilistic distribution; Aven (2016a) provides a full discussion.

repeatable:

> *"...risk analysis is scientific when understood as consisting primarily of (i) knowledge about risk-related phenomena, processes events etc. and (ii) concepts, theories, frameworks, approaches, principles, methods and models to understand, assess, characterize, communicate, and manage risk, in general and for specific applications"* (Hansson and Aven, 2014, p1181)

This segregation is useful to decision-makers in ensuring that important inherent properties of a risk analysis such as strength of knowledge, assumptions and uncertainties are communicated clearly. This is an improvement from risk estimations and quantified likelihoods of unknown accuracy, which create the potential for misunderstanding by the decision-maker. Delineating between value and objective elements of an analysis is highly relevant to describing the problems in quantifying structural integrity risk in operational situations because uncertainty frames the situation in which decision-makers are expected to make choices under time and resource constraints. The distinction suggests that a scientific process can indeed be found to contribute to risk research, and to allow incorporation of decision-maker values through execution of their judgement. Figures 1.2 and 2.2 are comparable in their separation of decision-maker judgement from the analytical phases of the risk calculation, validating the model from Hansson and Aven (2014) as representative of an industrial situation.

2.2.2.2 Uncertain about uncertainties

Risk analysis retains challenges to its scientific credibility, espoused by the lack of consensus in how to treat uncertainties (Hansson and Aven (2014), Shortridge et al. (2017)) and a lack of centralised risk policy across discipline boundaries (Aven, 2018a). This highlights the historical lack of curation of risk activities. Risk is more likely to be a secondary objective within primary research fields, as indeed might be considered in this study of aircraft structural integrity risk. Arguably, this has also hindered the consensus on the suitability of particular risk assessment

methods to specific situations (Hansson and Aven, 2014). To strengthen the scientific basis for risk analysis, knowledge creation surrounding situations of great uncertainty [16] is required to "conceptualize, assess, characterize, communicate, and manage risk" (Aven, 2018a).

The real-world issues facing risk assessment formalisations is their trustworthiness and appropriateness, particularly relevant in the building of 'safety case' to justify that equipment entering military service as 'acceptably safe'[17](Pasman and Rogers, 2018). Traditional/frequentist probabilistic risk assessments of unique situations with significant uncertainty (such as predicting the next aircraft structural failure or terrorist attack) add little value since the uncertainty cannot be reflected properly by probabilities and may mislead the decision-maker. Berner and Flage (2016) attempt to standardise approaches to uncertain assumptions by characterising situations using crude levels of belief in deviation from assumptions (low, moderate or high), and the sensitivity of a risk index (low, moderate of high). Each setting then guides the user to a particular uncertainty treatment method, depending upon the situation and on the strength of knowledge (strong or weak). Although this draws the use of assumptions and the treatment within a risk analysis to the attention of decision-makers (see also Aven and Zio (2014) and Aven (2016c)), the approach was only theoretically validated and does not account for rare events. Approaches to handling uncertainty in high consequence/low knowledge scenarios have been variously categorised: 1) Probabilistic Analysis, 2) Probability Bounds Analysis, 3) Imprecise Probability, 4) Random Sets, 5) Possibility Theory, and 6) Robust decision analysis, (Aven and Zio (2011), Shortridge et al. (2017)). However, as stated earlier, they are highly analytic and unsuitable for generalist aircraft maintenance decision-makers in time pressured, operational settings. Uncertainty treatment in such situations appears to depend upon subjectivity, judgement

[16] Situations of great uncertainty are also known as deep uncertainty, severe uncertainty, model uncertainty and wicked decision problems (Cox, 2012).

[17] A safety case is "a structured argument, supported by a body of evidence that provides a compelling, comprehensible, and valid case that a system is acceptably safe for a given application in a given operating environment" and may be a qualitative statement but targets stringent quantitative requirements (Pasman and Rogers, 2018).

and ultimately situational interpretation, which by Hansson and Aven (2014)'s perspective is a decision-maker's consideration.

2.2.2.3 Risk management strategies

How to apply a risk representation based on uncertain values must be considered carefully. Conservative strategies[18] are common in engineering, such as the use of safety factors in structural design (Elms, 1992, p28). They assure that risk tolerability criteria are not breached but they mask the magnitude of the true risk, may harm scientific credibility (Weiss, 2003), and should have the buy-in of decision-makers before being used (Berner and Flage, 2016). The cautionary (Aven, 2016b) and reasonableness principles (Weiss, 2003) might be considered sensible safety strategies for reducing risk, minimising the prospects of harmful activity exposure. From a temporal perspective, they both extend the time horizon of the decision, enabling knowledge building to reduce uncertainty and improve the management of the situation over time, termed "adaptive management" by Bodde et al. (2018). However, these approaches are unsuitable for the time constrained decision situation being faced by maintenance engineers.

Broader management of risk encompasses wider considerations beyond individual risk analyses. Some management strategies discussed in Aven (2016b) are relevant to aircraft maintenance structural integrity and described in the following.

Resilience The "ability [of an organisation or system] to sustain or restore its basic functionality following a stressor" (Aven, 2016b). This strategy frames the context of this research neatly. Common industrial risk management methods, such as fault trees and event trees, are inadequate for the needs of the structural integrity context. They are unable to predict unforeseen structural integrity events, learn from experience or respond to threats in a flexible manner - traits of a resilient system (Woods, 2017, p55-65). Traditional risk assessment methods

[18] Conservatism in risk assessment involves "a preference for erring on the side of overstating as opposed to understating risk under conditions of uncertainty" (Aven, 2016b)

thus appear unsuited to modelling complex systems like aircraft structure, being based on causal chains, historical failure data, event analysis, failure reporting, and are built by treating components with linear interactions (Dallat et al., 2019). Alternative methods are available such as Systems-Theoretic Accident Model and Processes (STAMP) (Leveson, 2004), which identifies controls and feedback loops that enforce safe operation, and determines which control failed for future accident prevention. But these methods are also time consuming to construct and resource intensive (Dallat et al., 2019). Integrating traditional and alternative methods is proposed to overcome these shortfalls (Aven, 2016b).

Robustness This is the "capacity for withstanding "vague approximations" and/or "zones of ignorance" in order to prevent undesirable impacts, notably the degradation of the properties to be maintained" (Aven, 2016b). A robust system requires no further risk analysis owing to inherent defences in a system to withstand a minimum level of adverse events. Justifying an assessment as robust necessitates quantification and sensitivity analyses of uncertain variables, which would be helpful for decision-makers but is unrealistic given time and expertise constraints for aircraft structural integrity maintenance risks. Inevitably, under operational constraints, being unable to declare a completed risk analysis as robust necessitates a decision-maker's judgement.

Criteria A set of *criteria* is essential for the decision-maker to demonstrate that a risk has met a pre-determined level. The UK Military Aviation Authority set a numerical threshold failure rates per flying hour (Military Aviation Authority, 2018b) and combine this with a cost-benefit analysis in to a subjective statement that a risk is As Low As Reasonably Practicable (ALARP). Although this allows comparison of risk option costs, setting criteria in operational circumstances is questionable where there are constraints on information and time. Uncertain probabilistic values may provide a baseline from which sensitivity analyses can be conducted to support decision-making, however this has already been discussed as significantly variable owing to the aleatory uncertainty in materials and usage (see Section

1.2). Maintenance managers are also not anticipated to be able to meaningfully apply such approaches in time constrained situations. Assessments of acceptable risk levels may also be influenced by framing effects, where individuals asked to assess probabilities relative to a scale may be influenced by negative or positive framing of the threshold (Yates and Stone, 1992, p61).

2.2.3 Risk Analysis Methodologies

Review papers provided signposting to relevant risk assessment methodologies, however each paper is structured relevant to their own objectives. No standardised taxonomy for categorising risk assessment methodologies was found. Aven and Heide (2009) classify methods as either relative-frequency-based, or Bayesian-based. Dallat et al. (2019) group methods under common methodological tites: Hazard and Operability (HAZOP) based; Failure Mode and Effects Analysis (FMEA) based; Qualitative and expert judgement-based; Tree analysis-based; Human Reliability Analysis and Human Error Identification-based (HRA/HRI); System-based; and Quantitative-based. Chemweno et al. (2018) reviewed quantitative methods specifically for how far they account for asset failure dependencies, grouping: Fault Tree Analysis; Bayesian Networks; and Stochastic Petri-Nets. Khan et al. (2015) reviewed risk management in process/petrochemical safety and categorised methods more generally by: Qualitative; Quantitative; Hybrid; and Semi-Quantitative. The Society for Risk Analysis summarise categories, methods and approaches that overlap the individual literature above (Society for Risk Analysis, 2018).

Many of these methods attend to *a priori* risk analysis, forecasting consequences and identifying appropriate mitigation activity before any hazard occurrence. However, the operational maintenance context concerns when damage has already (or is presumed to have) occurred, the equipment state is mid-deterioration to ultimate failure, and the risk assessment concerns the impacts of the damage on subsequent or future objectives and values. Recent discussions acknowledge the evolution of safety science towards ensuring that

systems are resilient to emergent risks (Swuste et al., 2020). High complexity and diverse risk sources are likely to characterise future systems, making it almost impossible to foresee all risks, necessitating dynamic, flexible interventions. The grouping by Dallat et al. (2019) provides sufficient resolution to discuss method relevance to aircraft maintenance decision-making, with the addition of *hybrid* and *semi-quantitative* methods (Khan et al., 2015).

2.2.3.1 Qualitative Methods

"Non-numerical representation and explanation based on attributes of graphics, flow diagrams, graphs and sources of data." (Khan et al., 2015, p120). Interviews, observational study methods, checklists, questionnaires, structured what-if analysis, expert elicitation methods such as the Delphi Technique, and scenario analysis are all types of qualitative approach (Dallat et al. (2019), Society for Risk Analysis (2018)). Checklists and observational study methods require self-inspection of situations, all others require a second party or team to elicit belief in potential outcomes. Checklists use established question sets to determine hazards across all aspects that might influence a process, although this is not relevant to assessing an emergent fault. Observational Study Methods indicate some relevance although post hoc in approach. In one example, clinical judgement was found to be as effective as formal risk assessment tools in the detection screening of patients for pressure-ulcers, generating a finding that resources may be better directed on risk management interventions than analysis (Webster et al., 2011). Qualitative expert judgement methods have advantages in that they are systematic in approach and application, cost effective, simple and user friendly. However they are prone to bias, dependent on the success of a quorum of experts, and do not necessarily consider systems issues (Dallat et al., 2019). Krause et al. (1995) advocated the use of argument to linguistically support qualitative risk assessment in situations where quantification is not possible. Their 'Standardised Argument Report' was applied to substance carcinogenicity, and presents a compelling case for the use of argument in risky situations where numerical assessment of uncertainties is unrealistic. However, the method applied

pre-established supporting and opposing arguments, which are not likely to be available to new and specific structural faults.

2.2.3.2 HAZOP Methods

Widely used in chemical and nuclear power industries (Pasman et al. (2017), Dallat et al. (2019)), HAZOP uses guide words to help brainstorming activities when identifying practices that may lead to risks within systems or processes. This method does not require identification of probability or severity and so does not quantify risk. HAZOP can be applied at any point in the design cycle, it is more 'upstream' of the context being investigated by handling the *identification* of risks. HAZOP is not suitable for this research context as it does not support event interactions easily, is time consuming to build and interpret, and is tailored toward the possibility of hazard occurrence - in a maintenance decision situation, the fault has already (or is suspected) to have occurred.

2.2.3.3 FMEA Methods

Developed to identify and analyse potential failure modes, effects and mitigation options by proactively questioning team members. The method identifies what might fail, what the failure mode might be and how frequently, and what the consequences of failure might be (Dallat et al., 2019). A number of principles may be useful to adopt when considering a structural integrity fault in aircraft maintenance such as the use of open questions and ease of adopting by practitioners. However, the full method is time consuming (Pasman et al., 2017), tedious and may not consider event interactions or human factors (Dallat et al., 2019). Anticipatory failure determination methods are suggested as qualitative alternatives (Thekdi and Aven, 2018), and although less rigorous than FMEA, is still not suitable for time-constrained situations.

2.2.3.4 Quantitative Methods

QRA "provides a realistic numerical estimate for better understanding and informed decision-making"(Khan et al., 2015, p127). This family of methods generates insight into where to best allocate resources, provides common understanding of uncertainties and dominant contributors within a system, and allows systematic identification of how a system can fail. Dynamic variants can improve resolution if updated through a project (Khan et al., 2015, p138). With regard to engineering risks, "the only QRA results that have a chance of influencing risk management are those that provide engineering insights"(Apostolakis, 2004), as opposed to financial or operations insights. While this "probably helps people design safer systems" Rae et al. (2014), the usefulness of QRA is coupled to whether decision processes are 'risk-informed' or 'risk-based'[19]. In the latter, the dependence upon validated QRA is likely to be of great importance but in the former (which is more likely (Apostolakis, 2004)), QRA is employed as one of many inputs in safety-related decision-making (Hansson and Aven, 2014). A prevalent QRA approach in the aviation industry is the risk matrix, which is advocated in the presentation, communication and management of risks (Figure 2.1). However, despite their simplicity, utility and appeal, risk matrices don't allow for aspects that are representative of real situations (Cox, 2008): the inclusion of budget considerations, risk reduction measures that can be applied, costs of risk countermeasures, risk propensity of the analyst, or whether the consequences or likelihood determined are influenced by judgement. These inherent problems with risk matrices challenge their suitability for risk management and decision-making.

Typically applied in tandem with other methods, QRA methods attempt to answer three questions (Dallat et al., 2019): what can go wrong? What are the consequences and how severe are they likely to be? How likely are these undesirable consequences to occur? In the aircraft maintenance context, the first question attends to the structural fault identified (or suspected), but the follow-on consequences given the emergence of the fault are uncertain. This is especially

[19] 'Risk-informed' is a position that risk assessment results are never the sole basis for decision-making by responsible groups, rather than 'risk-based' (Apostolakis, 2004).

so as there are limitations to what can be modelled (Apostolakis, 2004) and the foundational assumptions necessary for probabilistic algebra used in QRA methods may not be valid in real-world situations (Khakzad et al., 2011). Bayesian and other decision analysis approaches are suggested as possible solutions for uncertain and low data environments, which make it difficult to quantify system behaviour and model dependencies (Bier and Cox (2007), Aven (2016b)).

Bayesian approaches to QRA Complex systems with event dependencies, multiple variable states and parameter uncertainty, as found in aircraft structural systems, are suitable for modelling with Bayesian techniques (Langseth and Portinale (2007), Bier and Cox (2007)). They combine quantitative and qualitative sources of information through a systematic framework (Wilson and Fronczyk, 2016), which is useful and intuitive to use in low information environments (Chemweno et al. (2018), Dallat et al. (2019)). However, subjective influences such as the choice of method for combining evidence (Oberkampf and Helton, 2002) or the specification of priors (Bier and Cox, 2007) (see Equation 2.1) can bias the approach (Fischhoff, 1992, p146). Some claims, such as "where there is no observed data, the Bayesian approach has a reasonable solution" (Wilson and Fronczyk, 2016), lack real-world validation. Indeed a real-world representation with little information can only be validated when the true risk state is realised, which for structural faults is rare. Modelling probabilistic outputs through Bayesian approaches becomes increasingly more difficult and computationally demanding as the system grows (Wilson and Fronczyk, 2016). Thus, Bayesian approaches do not seem suitable for generalist engineering practitioners deliberating structural faults under the constraints experienced in operational settings.

Tree Methods Quantitative and diagramatic representations of a system in a tree-like structure are used to model the conditions that could result in an undesirable sequences of events, through logical gates representing faults. Event trees presuppose binary outcomes (success/failure) for events [20], and fault trees

[20] Dallat et al. (2019) considers tree methods separately from QRA methods because they include a pictorial representation of failure modes and knock-on impacts, alongside measurable values.

in particular have been widely used in the aerospace industry for safety studies (Khan et al., 2015, p132). More recently, the Reliability, Availability, Maintainability and Safety (RAMS) technique has been applied to complex situations such as airport security (Tamasi and Demichela, 2011), integrating FTA and HAZOP to assess barrier protection applying system failure probabilities. Although fault tree analyses allow useful diagnostic insight through the establishment of minimal cut sets[21], tree methods necessitate expertise and are time consuming to construct (Dallat et al., 2019). There is also the requirement to elicit subjective individual component probabilities of failure. Consequently, the quantitative aspects of tree methods are not expected to be useful for practitioners in operational settings deliberating structural failures. The qualitative diagrammatic form may be useful for visualising dependencies and causal structures.

2.2.3.5 Semi-quantitative analysis

Semi-quantitative analysis "falls between quantitative and qualitative analysis and produces approximate results rather than exact/absolute results"(Khan et al., 2015, p124) and "is useful when the direct measurement of process safety and risk is not possible whereas inference is acceptable"(Khan et al., 2015, p124). Using approximate inputs, semi-quantitative approaches appear to be more holistic (Khan et al., 2015) and may be suited to risk analysis situations where information and time constraints exist. In a structural integrity setting, gauging the suitability of inferences from historical data, past events or similar structural damage will rely on individual judgement, requiring the same degree of justification as described for subjectivity used within a quantitative analysis.

2.2.3.6 Hybrid analysis

Hybrid methods combine qualitative and quantitative methods but differ from semi-quantitative analysis because they provide "more precise and realistic quantitative results through its quantitative analysis"(Khan et al., 2015, p132).

[21] The minimal system component failures necessary to trigger top event failure

"Integrative thinking" in risk assessment is advocated by Aven (2016b) for creating a richer situational perspective. Although promising for use in process industry (Khan et al., 2015), extant hybrid methods appear unsuited to this research context. Stochastic petri-nets, for example, require significant computational power to solve rare failure events by simulation (Chemweno et al., 2018), and hybrid fault tree/Bayesian approaches suffer the same suitability concerns discussed individually already. The utility to structural integrity risk analysis in operational settings might combine the benefits of hybrid methods through a qualitative representation until measurable data is available, although this would only be applicable to repeatable or common failures.

2.2.3.7 Systems considerations

Accidents are largely a systems phenomenon (Rasmussen, 1997) so, by extension, identifying the extents of 'the system' being risk analysed helps reduce the propensity for accidents. Dallat et al. (2019) reviewed the extent to which current risk assessment methods conform to the 'tenets of accident causation' outlined by Rasmussen and whether wider organisational, managerial, social and cultural factors (which combine to form a 'systems' view) are taken into account. Linear, chain-of-event methods, like tree methods, that relied upon 'sharp-end' practitioners interpreting cause-effect relationships to represent the risk assessment scenario overlooked organisational and system influences. Meanwhile, bespoke methods that were designed to account for systems and organisational influences, such as STAMP described above, only made up 1% of the methods reviewed because they were found to be complex to learn and lengthy to apply (Dallat et al., 2019). The contrary effect is visible through the popularity of methods such as FTA, Event Trees and Bow Tie analysis in engineering maintenance environments. Their intuitive form and ease of use suggests that usability is key to engaging 'sharp-end' practitioners in meaningful risk assessment activity.

It is unsurprising that there is an increasing popularity of integrating methods to provide a 'tool kit approach' to risk assessment, since "no single approach

to risk assessment could be expected to suffice for all practical needs" (Dallat et al., 2019, p11). Methods that combine qualitative and quantitative features show the most promise in modelling real scenarios in the process industries (Khan et al., 2015). Overcoming the shortcomings of one method with the strengths of others may have practical difficulties however, requiring expert understanding of the methods, overcoming an initial burden to implement and use the methods, accommodating any computational dependencies, and adjusting the cultural norm to separate systems and business elements from 'mechanical' risk (Aven, 2016b). The discussion on hybrid methods above highlighted some of these.

Notably, Figure 2.2 (Hansson and Aven, 2014) is coherent with Dallat et al.'s findings concerning the extent to which systems considerations are manifest within risk assessment methods. In Hansson and Aven's model, wider systems effects are presented outwith the analytical fact-based elements of a risk analysis, contained instead within the decision-maker's value-based elements of a risk decision. This is sensible since factors like 'fatigue' or 'budgetary pressure' may not be modelled analytically but could influence how a decision-maker uses a risk analysis.

2.2.3.8 HRA/HEI Methods

Human factor focused methods are largely aimed at the assessment of human errors within safety systems at "the sharp end". They consider interactions between operators and the elements enabling their function but not the interactions between actors on different organisational levels, for example management influences on the risk assessment (Dallat et al., 2019). While these methods may model human effects within a system risk assessment, they have little direct relevance to the research context, since aircraft structural maintenance risk analysis considers physical artefact risk, not human performance. However, the substance of the methods directs attention to the human factors that are considered important in how humans handle risk, namely: the goals that humans desire and the physical and cognitive steps taken (task analysis); the thought processes underlying task performance (cognitive task analysis); the situation awareness of actors (situation

awareness methods); the team performance influences (team performance measures); and the cognitive workload of actors (mental workload methods) (Salmon et al., 2011). These parallel descriptive decision analysis work, discussed in Section 2.3.

2.2.4 Risk analysis quality

Understanding the effectiveness and usefulness of a risk assessment method requires a means of measurement, particularly given 'usefulness' claims by Goerlandt et al. (2017) and justification for QRA by Apostolakis (2004). Maturity models may provide an indication of QRA effectiveness and areas for improvement (Rae et al., 2014). However, scattered results across the 5 maturity levels proposed by Rae et al. (2014) suggest their model is inconsistent when applied to existing QRA methods, but may provide a useful checklist with which to prioritise improvement activity. An alternative method assesses the accuracy and usefulness of risk analysis simulations (Oberkampf and Helton, 2002), however simulation methods are not likely to be supportive for operationally time constrained engineering maintenance situations.

These two techniques sign-post the properties necessary for a good risk analysis, namely: being able to justify the accuracy and usefulness, that uncertainty must be understood clearly by stakeholders, and that risk analysis should be validated against the real-world (Oberkampf and Helton (2002), Rae et al. (2014)).

2.2.5 Applied structural integrity risk analysis

There is substantial literature regarding risk analysis as applied specifically to structural integrity problems. This section explores the current advancements in this area and is neatly framed by Figure 2.3, which highlights the different techniques applicable to the determination of a structural component's remaining useful life (RUL). All elements are covered within the following discussions except 'hybrid', which represents a collective methodological approach using several techniques to improve accuracy. The section is structured by the functional actions

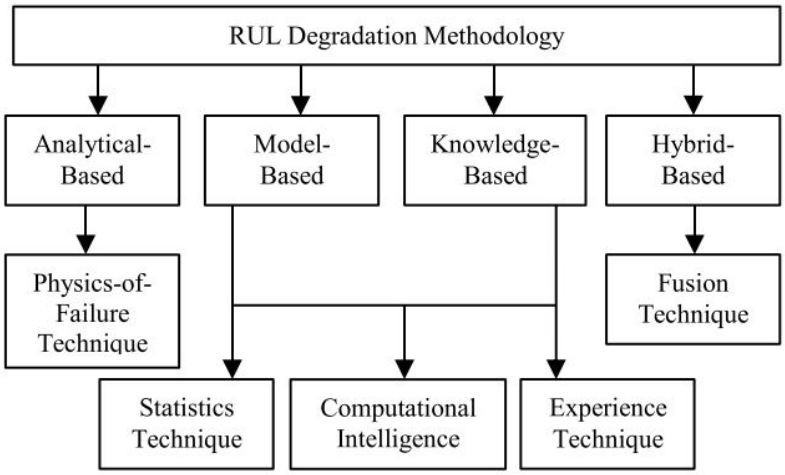


Figure 2.3: A classification of techniques for predicting structural failure and Remaining Useful Life (RUL), extracted from Okoh et al. (2014).

necessary for assessing structural damage: damage detection; and damage propagation. The types of damage considered are limited to the most common damage types: fatigue and corrosion, since progress in these areas is expected to indicate methods with suitable candidacy for addressing the research aim.

2.2.5.1 Damage detection modelling

Maintenance-based aircraft risks typically emerge following the identification of structural damage during servicing. Damage detection concepts have tended to centre on the direct detection of flaws in materials by the techniques outlined in Section 1.2.1. The growth of SHM as a comprehensive approach to handling the detection and management of structural damage is promising in forecasting and detecting damage initiation, although there are presently gaps between practical detection, analysis and management of structural risks (Frangopol, 2011). There are four key steps necessary for effective SHM, although no single method can comprehensively satisfy all of these steps (Vagnoli et al., 2018): 1) Fault Detection, 2) Fault Location, 3) Fault Severity, and 4) Remaining Useful Life (RUL) assessment.

In operational settings, analytical-based methods, such as finite element updating and knowledge hungry model-based methods such as artificial neural

networks are unsuitable for meeting the research aim. They are either too time consuming or relied upon training data being available, illustrated by Vagnoli et al. (2018, p998), which was heavily dependent on a "vast and reliable database of [civil] bridge behaviours". In aircraft, the structural configuration, historical loading and environmental exposure may differ across individual aircraft through life. Analytical parameters are unlikely to be meaningfully estimated by maintenance managers in low information environments, and irrespective of claims about successful applications of Bayesian networks to SHM, defining the prior probability distributions was one of the major issues influencing the efficacy of the method, which led to difficulties in assessing fault existence, location and magnitude (Vagnoli et al., 2018).

Situations lacking data have lead to criticism for analytical models, which were found to be over-conservative (Frangopol (2011), Yasseri and Mahani (2016)). Alternatively, Yasseri and Mahani (2016) apply a non-analytical 'importance factor' to specify a comparative model for safety (consequences of failure) of oil pipelines[22], and Frangopol (2011) compares the 'reliability importance' of a given bridge to the remaining bridges in a critical civil network of bridges. These approaches include structural calculations into an holistic view of the situation to allocate resources.

SHM helps improve knowledge about uncertainties in structural systems, although the relative importance of SHM information to a risk analysis depends upon the nature of the material being monitored and modelling assumptions made. For the fatigue case, usefulness depended upon whether the material was considered brittle or ductile, the fatigue damage threshold selected, the number of components in the system, the influence of damage on the failure of the system, and crucially the probability of component failure (Thöns, 2018). Although the findings, and the use of a utility network using a Daniels System[23] illustrates how SHM information may be integrated through a decision network, there is little

[22] Yasseri and Mahani conclude that the decision-maker must choose whether to improve upon the values found by their method, which is further justification for separating the risk analysis and decision-maker value-based judgements as argued by Hansson and Aven (2014).

[23] A Daniels System models structural redundancy assuming an equal load-sharing regime, and is derived as a statistical strength theory (Gollwitzer and Rackwitz, 1990).

relevance to the research objectives. This type of approach is more suitable to design contexts.

Damage detection and SHM methods rely on analytical methods for their predictions, and hence knowledge of the variables necessary for calculations. Although inappropriate in a time-sensitive maintenance environment, the principles of accounting for wider structural system considerations, like relative importance of components within a wider structural configuration may aid decision-makers deliberating the most important deterministic structural damage cues.

2.2.5.2 Damage propagation modelling

There has been much work on predicting damage propagation in structural materials, and the context of this work is framed by discussion in Section 1.2, notably that aleatory uncertainty is inherent in material crack growth (Virkler et al. (1979) and Wu and Ni (2007)). A majority of the works reviewed concern the handling of fatigue and corrosion damage as the two major contributors to structural failure (see Table 1.1).

Fatigue Damage Modelling Many of the modelling literature identified concentrate their arguments upon the choice of parameter distribution, their source of data, and the approach selected to propagate crack predictions temporally. Such approaches necessitate intimate knowledge of crack growth model parameters and material characteristics. Aircraft fleet maintenance has been modelled, using lognormal distribution to analytically model stochastic crack growth, real operating data, and subjective probabilities to calculate consequences for a damaged component (Yang and Manning, 1994). Fatigue crack growth has also been modelled using Bayesian approaches to update initial subjective priors with real measured data (Wang et al., 2009), although there was a lack of real-world validation.

Yan and Liu (2011, p162) clarifies the distribution choice a little further, that fatigue crack growth in “the fatigue of fighter plane [sic] is caused by maneuver load, and it follows lognormal distribution”, whereas “the fatigue of transport plane [sic]

is caused by gust load, and it follows two-parameter Weibull distribution". However, life predictions can only be made if assumptions regarding material properties remain consistent through-life. For instance, Yan and Liu do not account for damage encountered during service, use broad assumptions, safety factors, and assume that tests could be carried-out to identify model parameters. This approach is not suitable for maintenance situations where accurate model parameters are unexpected to be available.

Evidence Theory[24] has also been used to treat uncertainty in a crack propagation model (Tang et al., 2017). Using the Paris-Erdogan Law for fatigue crack growth modelling below, time-dependent factors are accounted for by treating parameters as parametrically dependent.

$$\frac{da}{dN} = C(\Delta K)^m \tag{2.2}$$

$\frac{da}{dN}$ is the crack propagation rate, C and m are constants that are dependent upon the material, environment and stress ratio, and ΔK is the range of stress intensity factor during a fatigue cycle given by:

$$\Delta K = K_{max} - K_{min}$$

Tang et al. (2017) treated all material variables as uncertain in an interval form with an assigned degree of belief for analysis by Evidence Theory, but depended upon using an experimental data set. Repeatable damage data is not expected to be available in all operational maintenance situations, particularly for new damage.

Since engineering maintenance practitioners are not expected to be materials experts and component material data is not readily available, the use of these methods for modelling damage propagation could result in inaccurate remaining useful life predictions in operational situations, unless bespoke tools complementary to the skill set of maintenance managers were accessible.

[24] Evidence Theory is also known as Demster-Shafer Theory and details a means of expressing one's *Degree of Belief* and combining evidence sources. It differs from probabilistic reasoning by applying probability values to *sets* of occurrences within a global set of possible occurrences, where as probabilistic reasoning applies probability values to individual events within a set. See He et al. (2018) for a clear explanation.

Bayesian Network approaches to crack propagation A number of authors have explored Bayesian approaches to structural integrity risk analysis, although not specifically applied to aviation. Straub in particular has been a strong advocate of BNs (Straub (2009), Straub and Der Kiureghian (2010) and Straub and Papaioannou (4134)), treating material deterioration as a stochastic Markov process[25]. A Dynamic Bayesian network (DBN) models a deteriorating structural system (Figure 2.4a), dependent upon time, time-invariant parameters, time-variant parameters, and initial damage condition, and is extended to include an inspection in Figure 2.4b. Using similar lognormal distribution assumptions about stochastic deterioration as in Yang and Manning (1994), the results conceptually demonstrate the structural modelling approach using Monte Carlo simulations, but are unproven against real-world risk systems. Variations of these approaches have been coupled to finite element analysis models to improve failure predictions (Li et al., 2017), and combined with Structural Reliability Methods[26] to improve modelling efficiency (Straub and Der Kiureghian, 2010). However, practical application of BNs to structural problems can be computationally intractable (Vagnoli et al., 2018), as well as time consuming to build, and requiring specialist understanding of probability and reliability analysis (Straub and Der Kiureghian, 2010). The concepts are highly analytic, are most likely beyond the capabilities of maintenance decision-makers, and would be beyond the decision horizons in operational settings to construct. However, noting that the use of BNs for causal modelling is more efficient than in diagnostic inference (Straub and Der Kiureghian, 2010), BNs may be helpful for follow-on risk analysis, as more data is collected.

Risk analysis for the damage tolerant philosophy Some literature specifically addressed the damage tolerant design and maintenance philosophy. Inspection data from routine damage tolerance maintenance inspections would

[25] A Markov process is one that assumes that the condition of the variable of interest at time t_1 depends only on the variable's condition at time t_0. Markov process modelling overcomes one key real-world modelling weakness of the Markovian assumption: that epistemic uncertainties do not generally vary with time.

[26] Structural Reliability Methods are used to compute the conditional probability tables (Straub and Der Kiureghian, 2010).

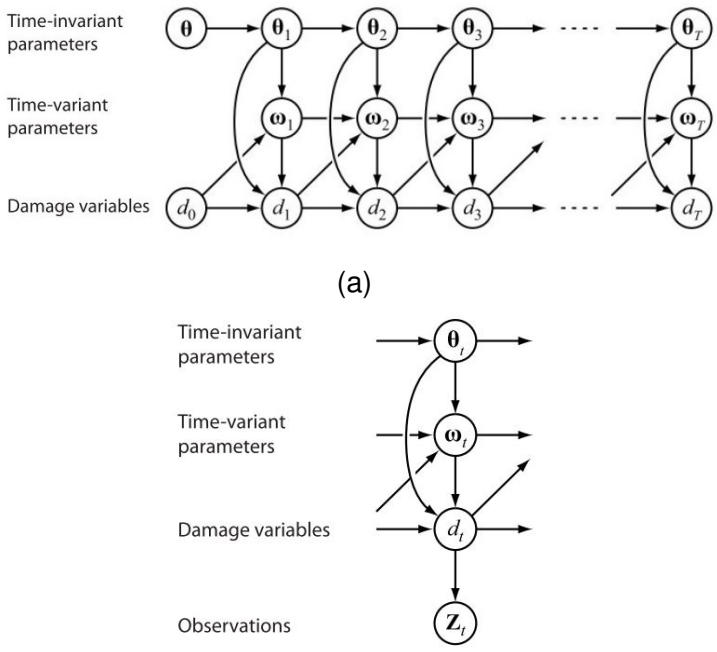


Figure 2.4: A generic Dynamic Bayesian Network for a deteriorating structural model in 2.4a and with inspections added as a child node in 2.4b, extracted from Straub (2009).

.

be useful when using a Bayesian approach to update failure predictions or in the evidence theory approach, but it then becomes essential for damage sizes to be accurately reported to enable a sufficient statistical body of evidence. Design assumptions must also be accurate to in-use condition, otherwise applying probabilistic methodologies is difficult (Rusk et al., 2001). Bayesian methods have been used to calculate the reliability of damage tolerance values using real data (Rusk et al., 2001), or by back-extrapolating crack sizes from up-to-date information on crack propagation in a method known as Equivalent Initial Flaw Size (EIFS) (Torregosa and Hu (2013), Cope and Moffett (2012)). This latter method exploits fracture mechanics relationships to establish a fictitious initial crack size, and extrapolate an expression for the Probability of Failure (PoF) (Torregosa and Hu, 2013):

$$PoF = \mathbf{Pr}(S \geq S_{cr})$$

where S represents the applied load, and the residual strength S_{cr} is given by:

$$S_{cr} = min(\frac{\sigma_{ys}}{\lambda}, \frac{K_c}{Y\sqrt{\pi a}})$$

σ_{ys} is the yield strength of the material, λ is a factor used to calculate the net section stress from the applied remote stress, K_c is the fracture toughness of the material, Y is a dimensionless parameter or function that depends on crack and specimen sizes, geometries and loading applied, and a is the crack length. EIFS uses the calculated (assumed) initial material surface condition of crack structures to calculate the crack length given the state of knowledge at that point in time. Cope and Moffett (2012) were able reproduce laboratory fatigue test results by using Bayesian methods to update EIFS calculations with SHM inspection data. The EIFS formulation is heavily influenced by the distribution of the crack size a at any given instance (Torregosa and Hu, 2013), by the quality and quantity of available fatigue crack growth data, and by being able to identify the geometry, loading and material variables (Fawaz, 2003), which may be different in-use compared to at design. Dependence on this expensive and difficult to obtain data (Wang et al. (2009), Virkler et al. (1979)) makes the method largely inapplicable to operational maintenance situations, where there is a paucity of data, structural expertise and time constraints on decisions. Furthermore, fleet-wide applicability of such analyses may be difficult to justify. Even similar aircraft being subjected to common loading spectra will exhibit differing structural damage characteristics through life, so a threshold for similarity must be understood before wider applicability can be declared. None of the methods reviewed provided any evidence of validation against real usage to substantiate claims, or took human influences into account, for example in the subjective judgements on statistical parameters or in the interpretation of SHM data.

Risk Analysis for the safe life philosophy Although superseded by damage tolerance and limit of validity philosophies (Section 1.2.2) safe lives are still relevant. In-service changes in aircraft utilisation and life cycle expectations challenge existing safe life limits (Hurtado and Hoffman, 2006), driving the

requirement to quantify the risk of fatigue. Bayesian approaches have been applied in the context of the safe life structural design philosophy. Given a specific damage location, Hurtado and Hoffman (2006) used the data from 25 aircraft to analytically estimate crack size probability density function at 100% FLE[27]. This approach depends upon knowledge of the critical crack length for the material (calculated from material fracture toughness), and would be an unhelpful approach for an newly emergent fault.

Suitability of structural design philosophy-focussed approaches Despite clear relevance to the research context, practitioners under time constraints would not realistically be able to construct a validated model of emergent damage in an operational setting, so these philosophy-focussed approaches are discounted as suitable candidates for the research objective. Both Cope and Moffett (2012) and Hurtado and Hoffman (2006) use Monte Carlo simulation to generate an initial crack growth sample and material properties, project crack size distribution for additional flying hours, requiring knowledge of material and geometric variables. Their success depends upon the construction of a valid computational simulation, which may not be timely or practicable. Additionally, only a single failure mechanism, strain-caused cracking, is assumed, which is overly simplistic in real-world situations where corrosive effects or other surface damages may be present. Combined damage mechanisms are considered further on.

Corrosion Damage Modelling BNs have also seen application in predicting corrosion initiation in concrete (Tran et al., 2016), although this particular reference only modelled corrosion initiation in a specific application, and was conceptual only using Markov Chain Monte Carlo simulation to generate random data to model. The lack of real data meant that the criticism of the method is limited to discursive narrative only.

[27] Fatigue Life Expended (FLE) is a US military measure of how far through the expected fatigue life a component is, where 100% FLE is the point when a full-scale fatigue test model (see Footnote 11) develops a crack of 0.01 inch.

Analytical-based methods were also found to be conservative when modelling corrosion deterioration in pipelines deterministically and probabilistically. Brown and Huyse (2012, p2) claimed that “the main objective of the deterministic assessment methods is to provide a minimum level of safety; accuracy of the prediction is really a secondary (although obviously desirable) concern”. Structural variables were also found to be time dependent in corrosion applications (Brown and Huyse, 2012), which is coherent with Torregosa and Hu (2013) and confirms that frequentist methods should be avoided for structural damage risk analysis.

2.2.5.3 Combined Damage Modelling

Modelling a scenario with both corrosion and fatigue is more realistic of real-world situations. Bayesian approaches have been applied to multiple material damage scenarios, with literature found across industries. Animah and Shafiee (2018) present a framework for integrated structural health decision-making to an oil and gas context, integrating condition assessment, RUL and life-extension decision-making. Notably, their paper shows coherence with the method categorisation used by Okoh et al. (2014) (Figure 2.3, Bayesian methods, statistical techniques and stochastic methods, computational intelligence, physics of failure, and expert elicitation), and in the four-step SHM process used by Vagnoli et al. (2018) (fault detection, fault location, fault severity, and RUL assessment). These commonalities illustrate the cross-industry experience of the same methodological problems when handling structural integrity risk. For instance both aviation structural damage and oil and gas contexts lack reliable data or training data, meaning computational intelligence methods may be impossible to use, and expert elicitation becomes necessary (Animah and Shafiee, 2018).

Chen et al. (2018) model the combined effects of fatigue damage and corrosion growth in an oil and gas steel catenary riser using a BN based upon a probabilistic solution of the Paris Crack Growth law (Equation 2.2). Cracking was dependent upon the corrosion model, which used a subjective probabilistic wall thinning rate per year on the exterior of the structure as a parent node to the experienced stress range (part of the cracking model). This approach offered no real-world

validation of subjective choices, such as wall-thinning rate in the corrosion model, or the maximum likelihood estimate for fatigue growth. Ma et al. (2014) demonstrated improved accuracy of a Bayesian network-based model by updating with real-world bridge deflection data. But the iterative approach and dependence upon computational simulation to produce posterior distributions is not suited to operational settings.

There are few examples of where judgement ('experience technique' in Figure 2.3) has been applied to real scenarios to model structural damage. Relative importance has been used to predict the consequences of a risk (Frangopol, 2011), and expert elicitation of the relative importance of railway bridge components was used in a structural deterioration 'what if' scenario analysis in combination with a DBN (Rafiq et al., 2015). The latter approach used only visual inspections of component conditions as the only input to the model, which suggests utility to maintenance situations. However, the method necessitates a configuration-specific model, and in aircraft structural maintenance scenarios with infrequent failure data, would be dependent upon maintenance managers to construct a DBN. This is an unrealistic expectation under operational constraints.

2.2.5.4 Structural integrity risk analysis summary

The difficulty of applying appropriate management methods to handle material damage identified in-service is not unique to aviation, with supporting literature from oil and gas (Animah and Shafiee, 2018) and rail (Vagnoli et al., 2018) domains. Predicting material failure in situations of risk is time consuming, requires significant expertise to configure, populate and analyse models, is information hungry, computationally demanding, and does not presently have a recognised best practice. Although some literature used real damage measurements to generate a risk analysis, the results were insufficiently validated against true risk, so the claimed benefits cannot be verified. Practical realities of applying the methods discussed to operational aircraft maintenance situations have been highlighted. The complexity of the structural context necessitates deterministic models that can only simulate fracture growth using computer-based algorithmic solutions. All

models required fracture mechanics variables to be known or rigorously simulated to provide analytical accuracy, the data for which may not be available in an operationally time essential situation. Relative indices provided an alternative representation of the structural integrity context but were not validated, relied upon elicitation of parameters from experts, and were only useful for identifying methodological approaches.

Given the requirements outlined at the top of this Chapter, the methods of calculating structural risk, by either corrosion or fatigue material deterioration mechanisms, are not suitable for the situation being studied. Although there was widespread acknowledgement of the aleatory uncertainty faced when modelling structural failure, the methodologies reviewed in this section did not consider analyst capability nor the *operational* situation, where information, time and resources may be limited. The methods reviewed in this section would be suited to equipment manufacturers with specialist knowledge, who are away from the operating maintenance environment and provide 'reach-back' support for UK military aviation maintenance teams.

2.2.6 Relevant non-structural applied risk modelling

A wider net was cast to identify relevant approaches from risk situations in operational circumstances, when information is poor, and where there is particular consideration of the real-world situation, outside of a purely theoretical and laboratory research setting.

In hazardous maritime scenarios, BN representations can be updated with new information as it incrementally becomes available (Eleye-Datubo et al., 2006). Although input values taken from experts or subjective distributions were not justified in this example, the visual format of a BN was claimed to ease the communication of causal chains to layperson stakeholders. This echoes earlier discussions of BNs regarding the transparency of BN parameter choices for decision-making.

The intelligence and security sector shares similarities to the structural integrity

maintenance context: uncertain scenarios; insufficient resources to acquire knowledge regarding variables in the risk analysis; and time constrained decision horizon (Isaksen and McNaught, 2019). The presence of intelligent actors with intent effecting events in security settings is a crucial difference - engineering factors contain no behavioural influence on hazard initiation. Notwithstanding, the discussion on rare event analysis and defining the strength of knowledge of assessments in Askeland et al. (2017) highlights that the use of historical data in rare event analysis is not necessarily reliable. Where there is no 'stability' to rare occurences, frequentist statistical analysis cannot be applied. Similarly, the lack of transparency in the definition of uncertainty reduces the quality of risk analysis and risk communication (Isaksen and McNaught, 2019). With these commonalities in mind, risk analysis in operational maintenance environments and the security sector stand to gain from improvements in handling risky decisions, where there is insufficient information available for a deterministic appraisal.

2.2.7 Section Summary

The review of risk analysis state of the art has identified common shortfalls across disciplines: consistency of approach; handling uncertainty; separating fact-based analysis from value influences; and the validity, reliability and accuracy of risk analysis. In regards the treatment and handling of uncertainty given poor data, knowledge or information, there have been incremental improvements made to both the description of uncertainty, through qualitative (strength of knowledge statements) and quantitative (probability bounds, interval analysis) means, and in the handling of uncertainty through dynamic modelling (dynamic Bayesian networks, dynamic fault trees), hybrid representations (hybrid fault tree/bayesian representation), and simulation strategies (Markov Chain Monte Carlo).

Considering methods that handle aircraft structural integrity specifically, a majority of the methods reviewed were hungry for information that would not be readily available in maintenance environments. Indeed, making subjective judgements on material parameters in the field may be considered too crude for

the fine-grained analysis techniques typical of material performance calculations (Pitz, 1992, p315). The analysis methods that claimed apparent validity by applying real-world case studies were beyond the expected capabilities of maintenance decision-makers, and required time to either run simulations or to elicit subjective data from experts. This is not to imply that the methods explored are not relevant, only that they are not applicable to the constrained settings that this research is targeting. Although advances in structural health monitoring will mature in due course to provide real-time data for analysis, current aircraft maintenance environments still rely upon human decision-making and will continue to do so until trust in automated health monitoring is gained.

Researchers have expressed the need to include decision-makers in methods for handling uncertainty, acknowledging the need for assumptions to be communicated and understood. After all, the decision-maker will be the main benefactor of a risk analysis as a consideration in their final judgement. This inevitably makes risk analyses about rare events or situations where accurate values are unobtainable a more integrated problem that exceeds the bounds of traditional risk analysis formulations. This highlights the need to include consideration of human influence on decisions about risk. No effective guidance cognisant of decision-maker situational capabilities for handling structural integrity risks was identified.

2.3 Decision analysis domain review

Decision-making is the process of making a choice from a set of alternatives, the outcome from which depends on the state of the world (Peterson, 2009). Different paradigms of decision-making research remark upon distinct ontological perspectives and epistemological facets, which correspond to how an individual might approach a decision-making situation. Howard and Abbas (2016, p28) categorise types of decisions in a hierarchy with a combined time domain and resource allocation axis in Figure 2.5. This representation does not account for the situation being faced by this research project - decisions of great importance

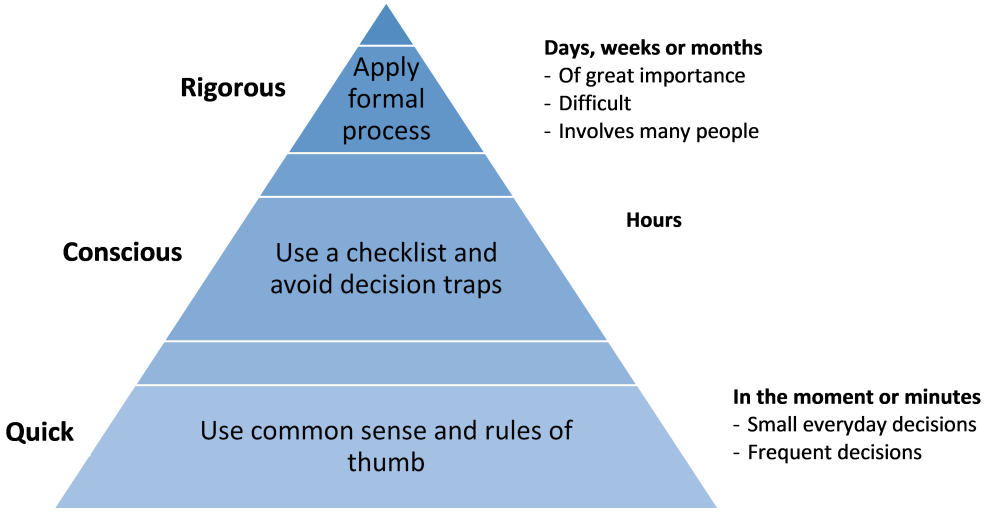


Figure 2.5: Hierarchy of Decisions, extracted from Howard and Abbas (2016).

concerning the safety and airworthiness of aircraft, that are difficult to analyse but are required to be made in short periods of time with few resources. Decision approaches are reviewed to identify opportunities to support the research aim.

2.3.1 Decision analysis approaches

There are three primary approaches to Decision Analysis that describe their ontological perspective of decisions:

– The *normative* approach seeks to explain *how* decisions *should* be made, if they are to be made *rationally*, developed by mathematicians and economists (Hoffrage and Marewski, 2015). This decision theory defines logical axioms (or a 'norm') for the rational actor, given the information available, and methods developed under this approach conform to these axioms.

– The *descriptive* approach seeks to explain *how* decision-makers *actually* act when making decisions (Peterson, 2009), developed by psychologists (Hoffrage and Marewski, 2015). This approach details that cognitive and motivational biases cause distortions of judgement because of internal influences (cognitive) or desires for specific outcomes, events or actions (motivational) (Montibeller, 2020).

– The *prescriptive* approach describes a combination of normative and descriptive decision analysis, and is also termed decision aiding. It is an applied decision analysis paradigm configured for real-world scenarios. A decision aid should be "interdisciplinary...logically sound, cognitively realistic, organisationally appropriate and take advantage of the best practical skill and knowledge available in an application domain", in order to be used and useful (Brown, 2015, p213). This approach bridges the gap between the *normative* and *descriptive* approaches, and real-world complex decision-making situations.

Normative decision analysis (what *should* be done) corresponds to the deliberative "cool cognitive" system that leads to rational behaviour; and descriptive decision analysis (how people actually make decisions) studies the affective "hot emotional" system that is motivated by fear, sex, hunger and other basic stimuli related to early human survival (Howard and Abbas, 2016, p23-26) that cause deviations from rational action. These two approaches have varying descriptive titles across the literature reviewed, and have been described as Type or System 1 (descriptive, animalistic) and Type or System 2 (normative, analytical, using the pre-frontal cortex) types of thinking (Hoffrage and Marewski (2015), Aven (2018a), Evans (2019)), Table 2.1 provides a summary. However, they are typically studied as separate fields. For example a normative analysis of casino games would identify the rational reasons why we shouldn't gamble, however descriptive analysis can identify why some people still decide to.

The *prescriptive* methods have emerged more recently from desires to provide supportive research to applied real-world scenarios (Klein, 2008). Which approach should be followed in an engineering situation remains debated (Crossley et al., 2017). A culture of objectivity appears to have prevailed in applied situations because there is insufficient data to support judgement-based approaches and because of perspectives such as: "subjectivity and feelings are private business" (Hoffrage and Marewski, 2015). Decisions are hence individualistic and automatic, using decision short cuts (heuristics), guided by gut feelings, or applied judgements. Although efficient in many circumstances, heuristic approaches are vulnerable to

biases (which is the focus of descriptive decision research) or information that later turns out to be false. Given the information and time available to make a decision, knowing when to make a heuristic-based decision, or choosing to source more information and follow a more normative approach before making a decision is crucial (Crossley et al., 2017). Further complications are added when situations are constrained by the available information, time and resources for an engineering decision, restricting the extents to which the situation assessment and decision choices can be fully explored. In Figure 1.2, sources of motivational or cognitive bias, such as how line management interpersonal relationship may influence a decision judgement, were not shown. Ultimately a choice is made, which is influenced by the judgement and cognitive ability of the decision-maker.

Conducting a risk analysis would be part of this culture of objectivity, where a rational decision-maker would be anticipated to act *rationally* on the analysis results, usually selecting the action with the least risk calculated for it. Using risk analysis results as a component in a broader decision consideration inevitably means that the personal *objectives* and *values* of importance to the decision-maker have influence - what variables should be included in analysis? What measures are used to define outcomes? What are the thresholds for action on these measures? Some identify that the 'art' of risk analysis is the selection of appropriate models and parameters to accurately describe the real-world situation being assessed (Theofanous, 2003). These subjective considerations are important to understand when using the outputs from a risk analysis model. However, in the aircraft maintenance context, structural integrity parameters are either unavailable or too cumbersome to determine in the time available and any risk analysis output generated becomes strongly influenced by decision-maker judgement. How great this influence is depends on the models used in organisational decision-making.

The importance of human cognition as a factor in engineering decision-making requires greater exploration (Vermillion et al., 2015). While approaches, such as Bayesian techniques, attempt to overcome the limitations of subjective situations, their apparent rationality is still influenced by a decision-makers cognitive or motivational biases, for example in the selection of uninformative prior parameters

Normative	Descriptive
System 2	System 1
Objective	Subjective
Cold rational	Hot emotional
Analytical	Intuition
Deliberative	Automatic
Verbal	Experiential
	Animalistic
	Narrative
	Non-verbal
	Natural

Table 2.1: Comparison of the lexicon used in literature to describe Descriptive and Normative Decisions.

(Fischhoff (1992, p146), Hoffrage and Marewski (2015)). Other factors that are beyond the immediate decision may also influence a maintenance manager's decision. Visualising the decision sequence for a particular circumstance may include:

– Other equipment may be effected (for example, other aircraft in the same fleet) by precedent.

– The maintenance decision may have knock-on impacts to other maintenance activity, scheduling and resourcing.

– The initial decision may require review following new information outside of the time horizon of the initial decision.

– The initial decision may need to be reviewed to understand why a particular decision was taken, for instance during a post-incident inquiry.

Combined with the information available to act upon the immediate decision, these situational factors may bear upon the decision-makers' preferences and perspectives on rationality of the decision basis.

2.3.1.1 Rationality

The aim of decision-making is to make choices that maximise one's goals. When our mental states or processes help us achieve this, those states and processes are considered *rational* (Over, 2005). Perceptions on rationality vary, depending on the epistemology of ones focus. For instance, normative decision advocates believe that rationality is independent of the individual and normative decision theories should "yield prescriptions about what decision-makers are rationally required to do" (Peterson, 2009, p3) based upon the available information. This perspective aims to establish an optimal decision solution. However, whether those prescriptions are suited in a particular context for achieving one's particular goals, is what leads some philosophers to consider rational standards, and thus formal decision theory, as too forceful (Over, 2005, p6). *Bounded rationality* provides a competing view, which in a decision-making context means that individuals can be satisficed[28] by a particular choice, while falling short of the optimal (Simon, 1955). This posit makes greater allowance for what is possible in a given situation, as limited by the conditions of the circumstance and the ability of the individual.

The relationship between rationality and aircraft maintenance structural fault decision-making is thus: there are normative tools available for analytically deriving structural integrity risk, but these optimal solutions are out of reach for maintenance managers in time-constrained situations. Maintenance managers make their choice from a perspective of bounded rationality: within their cognitive and situational limits. This competes with a global perspective of rationality, where all available choices can be appropriately assessed. To support maintenance decision-makers using their judgement, prescription of an approach must be procedurally rational, which replaces "the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by humans" (Simon, 1955, p99). Supporting maintenance managers compelled to make decisions about risk

[28] A satisficed decision produces a choice that is satisfactory or good enough, but not necessarily optimal. Gainsburg et al. (2016, p334) discusses how this is often the objective for engineers, who may specify their engineering best answer to a problem, although it is never *the* answer, unlike in scientific conflicts where there must be a clear solution.

in constrained situations thus attends to *how* the decision is made (procedural rationality)(American Psychological Association, 2020a), rather than *what* decision is made (substantive rationality)(American Psychological Association, 2020b).

2.3.1.2 Utility

A unifying theory about rational choice regards utility. During selection of a choice (and the subsequent follow-on considerations) an actor will have an individual preference for particular outcomes. The concept of *utility* creates a relative scale for the decision-maker and is an important generalised measure in decision theory. Taking Figure 1.2 as an example, one may define ordinal preferences for maintenance intervention as:

$$\text{Defer maintenance } (U_1) \succ \text{Replace component } (U_2) \succ \text{Repair immediately } (U_3)$$

The term $\succ$ represents preference on a scale of "goodness" (Baron, 2005, p23-27), where 'Defer maintenance' is preferred to the 'Replace component', which is preferred to 'Repair immediately'. Assignment of quantitative values to these preferences allows relative comparisons between variables on an interval scale. Expected Utility Theory[29] (Savage (1954)) describes decisions under uncertainty (Baron, 2005, p24) and is generalised to account for individualistic probability perspectives as Subjective Expected Utility:

$$\text{SEU}_i = \sum_i u(x_i)P(x_i)$$

Where $u(x_i)$ represents a decision-makers 'utility function' or value measure for the outcome events x_i; and $P(x_i)$ represents the probability of the events x_i derived subjectively or otherwise. This relationship links utility and probability together, such that if a decision-maker is rational and believes that the outcomes x_i are possible, with corresponding utility $u(x_i)$, then the decision-maker's choice can be explained by their utility of the outcome combined with their belief in the

[29]The theory built upon the work by von Neumann and Morgenstern (1947) and Ramsey (1926).

probability of the outcome $P(x_i)$. Although the approach enables the combination of preference with individual values to optimise decision-making, it suffers the same issues as QRA, namely that the perceptions of an event's likelihood are encoded into probabilistic estimates and hidden from scrutiny. A diagrammatic extension of utility theory is commonly seen throughout the literature, called *influence diagrams*, which link 'chance' nodes/events (X_i) with 'decision' nodes/actions/interventions and 'utility' nodes. If there are n number of X nodes, 'maximising the utility' represents the node and decision combination with the highest expected utility. Although they help modellers visualise a decision problem, they are not suitable for operational situations where validating a requisite diagrammatic model would be unlikely, not withstanding the need for probabilistic information.

2.3.2 Elements of a decision

There is a lack of consistency in defining what a decision consists of. Even across the sample of books and papers from the countless options that provide a description of a decision, there is varied lexicon and the framing of analysis as either a *staged methodology* or describing the *elements of a decision*:

Clemen and Reilly (2001) - (*elements of a decision*): values and objectives; decisions/choices to be made; uncertain events; consequences.

Clemen and Reilly (2001, p5) - (*staged methodology*): Identify the decision situation and understand objectives; identify alternatives; decompose and model the problem (problem structure, uncertainty, preferences); choose the best alternative; sensitivity analysis; identify if further analysis is needed; implement chosen alternative.

Mellors and Locke (2007) - (*staged methodology*): define the problem and set goals; gather information and identify options; evaluate information and options; make choice; implement choice and monitor results.

Peterson (2009) - (*elements of a decision*): states; acts; outcomes.

Yates et al. (2011) - Cardinal Issue Perspective (*elements of a decision*): Preliminaries (*Need, Mode, Investment*); Core (*Options, Possibilities, Judgements, Value, Tradeoffs*); Aftermath (*Acceptability, Implementation*).

Although each example uses differing terminology, the descriptions follow a broadly similar process and can be arranged in the following amalgamation:

- There is a decision-maker in an environment that has a defined set of values and objectives.
- There is a (are) problem(s) that requires intervention, which can be imagined with data, information or knowledge available to describe the potential future outcomes for the hazards.
- There are identifiable actions available for the decision-maker to select to mitigate the effects of the hazard, based upon the information available (including broader relevant situational knowledge), the decision-maker's judgement, values and objectives, and the consequences that may be felt from the action.
- Outcomes of the action chosen should be monitored for effectiveness in implementation.

This process parallels risk assessment in which values are not included: identify hazards, perhaps representing these through a fault or event tree; then structure the risk quantification representation by modelling the system being analysed, the environment the system is operating within, and a decision-rule that describes the threshold for intervention or action (Bier and Cox, 2007). Even when informed by probabilistic risk analyses, many decisions are made without formal rules, either because the risk assessment results make the best decision obvious or because of the need to address wider concerns of other stakeholders (Bier and Cox, 2007). This emphasises the key distinction between risk analysis and decision analysis: the inclusion of values, objectives and subjectivity of the decision-making elements. This was observed earlier in Figure 2.2 and corresponds to the limitations of risk assessment methods explored in Section 2.2.2. Rare events make QRA difficult to

complete or misleading, but the subsequent decisions are predominently based upon human intervention and therefore subject to human fallibilities. There are few articles that address bias in risk management and decision-making, despite the obvious relevance (Montibeller and von Winterfeldt, 2015). In situations where accountability requires transparency, being able to justify a decision basis is important, underscored by the research requirements (see item 5 at the start of this chapter). Decision analysis addresses this across the three paradigms of normative, descriptive and prescriptive decision analysis but across differing ontological perspectives of a decision problem. The differences in perspective can be viewed as a difference between how a human behaves in decision situation, and what choice should have been selected in hindsight.

2.3.3 Normative decision analysis

Traditional decision analysis and theory is centred around the quality of rationality. Normative decision analysis is built upon the fact that irrespective of the community of decision-makers, the organisational structure or culture, and the methods used - rationality, as a thing, is not expected to change. On this basis, normative decision analyses attend to axioms for rational action that have long been established (von Neumann and Morgenstern (1947), Savage (1954)) and challenged (Clemen and Reilly, 2001, p578-582). Techniques developed to support rational decision-making rely upon the assumption that decision-makers are rational actors and will select the decision alternative that optimises their goal (e.g. money, utility, probability), known as substantive rationality (Simon, 1979, p67-68). There are numerous techniques that use these two assumptions as a basis for their scientific foundation, such as Maximum Expected Utility, Cost Benefit Analysis, Dominance, Maximin, Maximax, to name but a few (see Peterson (2009) or Clemen and Reilly (2001)).

Normative decision analysis techniques can be used in situations where probabilities of states are known (defined as ‘decisions under risk’ Peterson (2009)), or where outcomes and decision alternatives are known but probabilities

are unknown (termed 'decisions under ignorance' Peterson (2009)[30]). Where probabilistic interpretations are required, probabilities tend to be elicited from experts if they are unavailable from system data. Expert elicitation is itself a research field, with methods such as the Delphi Method (Merkhofer, 1987, p747) developed to minimise individual and team bias through comprehensive frameworks, ultimately aiming to achieve consensus across the experts involved in the elicitation process. Typically, such processes require a number of experts over a number of days to complete the process from identifying the objectives and system representation through to agreement on assigned values.

Some normative decision approaches are already familiar. Bayesian Networks (seen in a decision context earlier in Thöns (2018)) and Event Trees[31] are developed further to include measures that structure preference (Howard and Abbas, 2016, p123), such as monetary measures (Jordaan, 2005, p170), or the generalised measure of 'utility' (see Section 2.3.1.2), in order to provide decision support. Nielsen and Sørensen (2010) physically model structural deterioration in offshore wind turbines by the decision analytic extension of BNs, influence diagrams, using similar BN representations as described in Section 2.2.5.2 by Straub and Der Kiureghian (2010). As highlighted previously, such analytical representations are computationally intensive and require knowledge of the structural parameters, which therefore renders them unsuitable (see Section 2.2.5.2). Pate-Cornell (2007) use an Influence Diagram to structure risk analysis in their SAM model[32], accounting for probabilistic characterisation, actions and management factors through a diagram to benefit communication. Using available evidence (as opposed to frequentist-type data, available or simulated), they argue that their evidence-based decision-making process is rigorous, despite inputs being subjective or imperfect. Guthrie and Reid (2018) present an event tree for holistically modelling pipeline failure induced by landslide. Populated with expert judgement and known information, the event tree branches are organised

[30] The term 'decisions under ignorance' is less relevant with the wider acceptance of two degrees of uncertainty - aleatory, and epistemic discussed in Section 2.2.1.2 - which grant greater clarity to an uncertain situation and may affect the decision

[31] Event Trees are also described as 'probability trees' in Howard and Abbas (2016)

[32] The SAM model was first presented in Murphy and Pate-Cornell (1996).

so that decision nodes are grouped across branch paths, to allow comparison across options. The potential for inaccuracy caused by influences upon subjective judgements is acknowledged, but Guthrie and Reid (2018) and Pate-Cornell (2007) lack discussion on the impact this may have to the analytical outputs and the decision-maker's selection. Decisions based upon judgements must be justified to enable transparency, reassure stakeholders, and enable accountability, but are unable to validated until sufficient evidence has been gathered (Pate-Cornell, 2007).

Ecology settings provide similar information constraints to operational maintenance structural fault decisions, where it is prohibitively difficult to obtain information from experimentation. Structured elicitation techniques have been combined with BN (Barons et al., 2018) and influence diagrams (Thorne et al., 2015) in streamlined approaches, using consensus to provide transparent and defensible evidence within a complex system model.

When considering the relevance of applying normative techniques to support decisions regarding structural risks that emerge during aircraft maintenance, frequentist probabilities of events are not appropriate and subjective probabilities must be provided by the decision-maker without expert support. Time constraints are also likely to be too restrictive for a meaningful normative model to be constructed. Normative methods, seeking substantive rationality about the decision outputs (Simon, 1979), offer rigour and a vast body of knowledge for handling situations in which the time available is greater than or equal to the time required for the decision problem, but this paradigm only considers compatible situations and exposes a shortfall in decision research described by Figure 2.6. Under traditional normative approaches, complex decisions would be assessed optimally with sufficient time to build a requisite model and analyse the situation, as represented by points A, B and C. However, there is insufficient literature considering appropriate techniques for complex decisions required under time constraints, as at point D. Similar observations were made by Simon (1979), who noted that classical economic analysis pays inadequate attention to how human beings behave in the face of uncertainty, and advocated theories of procedural

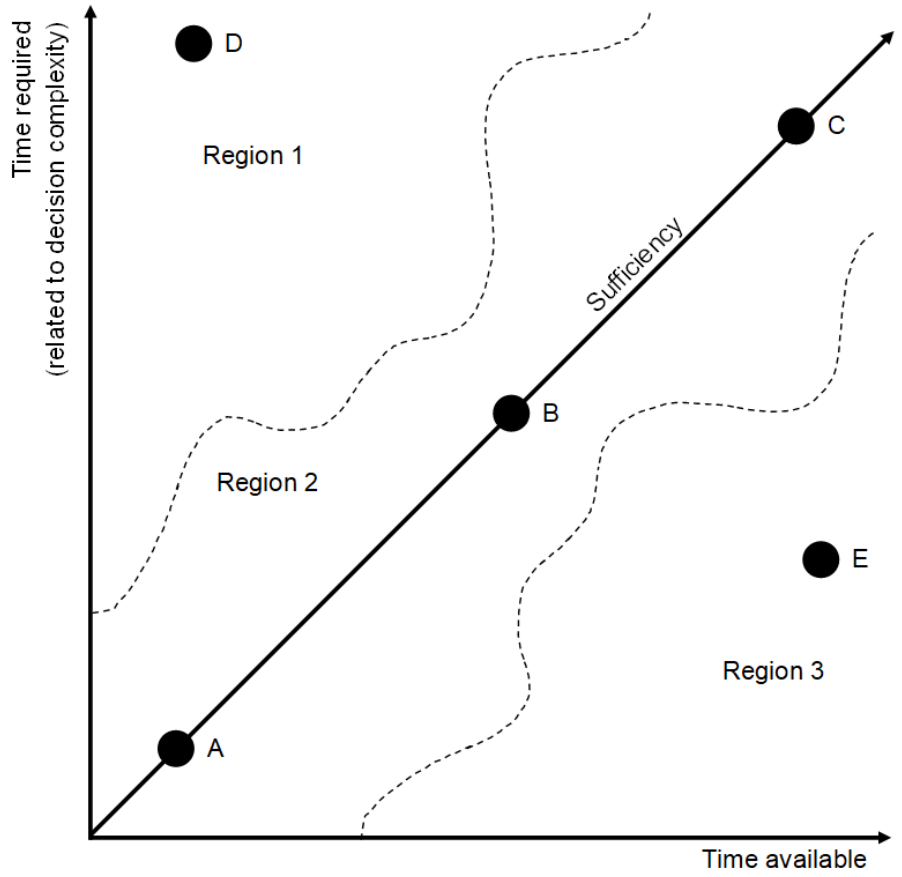


Figure 2.6: An extended decision-making hierarchy, adapted from (Howard and Abbas, 2016), showing an optimal locus, the line of sufficiency, flanked by sub-optimal regions that may be constrained by the context or influenced by decision-maker characteristics.

rationality that are consistent with real-world dynamics.

A further complication concerns accountability in situations with poor knowledge. Accountable decision-makers must be able to support their choices with a means of justification to enable stakeholders to understand assumptions, the knowledge base and potential sources for bias. The use of argumentation to aid justification in normative decisions has been seen applied to medical toxicology decisions (Krause et al., 1995), design decision-making environment (Polacsek et al., 2018), and in discussion about justifying elements of risk analysis (Hansson and Hirsch Hadorn, 2018), but no application has been identified to practical engineering maintenance situations.

2.3.4 Descriptive decision analysis

The descriptive decision analysis field is vast, covering sub-divisions of the field such as the Information Processing Approach, Prospect Theory, and Behaviour and Actions (Beresford and Sloper, 2008). This section provides insight into research that has explored how decision-makers are influenced, to help identify whether decision-makers that use judgement can be better supported. No literature has been identified that specifically addresses descriptive decision analysis with respect to aircraft structural integrity decisions.

With it's roots in psychological sciences, identifying *how* and *why* decision-makers select the choices in the way they do has established usable knowledge on the sources of influence, based upon empirical evidence. Early works on Prospect Theory (Kahneman and Tversky, 1979) counterbalanced the work on rationality described in Section 2.3.3 by identifying that decision-makers are not always rational and are influenced by heuristics, bias and framing, which leads to distortions in normative decision-making. These influences may be the consequence of decision-makers emotion regarding: the decision subject matter; the process of making a decision; or regarding the aftermath. They are either cognitive biases (conscious/subconscious distortions of judgement), or motivational biases (distortions owing to self-interest, social pressures or organizational context) (Montibeller and von Winterfeldt, 2015). They can, for instance, lead certain decision-makers with a risk-avoident propensity to require more information before making a decision (Buchak, 2010), or they may mislead stakeholders with overly optimistic forecasts for a preferred course of action (Montibeller and von Winterfeldt, 2015). Individual appraisals of decision situations are automatic (hence termed 'hot/emotional' responses) and research into such reactions must be conducted retrospectively (Salmon et al., 2011). Well documented heuristic responses are (Tversky and Kahneman (1974),Slovic et al. (2004),Bazerman and Moore (2009)):

– *Availability Heuristic.* Enables judgements by relying upon immediate examples that come to mind. Situations may be biased by experiences that

have recency and vividness (their ease of recall), how memory structures are established (retrievability), and any presumed associations

– *Representativeness Heuristic.* Enables judgements on the similarity of objects of people to an existing persona, and can be influenced by insensitivity to base rates, sample size, misconceptions about chance, regression to the mean, and the conjunction fallacy[33].

– *Anchoring and Adjustment.* Uses cues to guide judgements, however people have been found to be poorly guided by cues, even if the relevance of the cue is unrepresentative of the decision situation. Typical biases are insufficient anchor adjustment, conjunctive and disjunctive events bias, and overconfidence.

– *Affect Heuristic.* This refers to an automatic emotionally judged response, a specific quantity of “goodness” or “badness” (Slovic et al., 2004), such as perceiving human-caused hazards as more severe than naturally-caused hazards (Siegrist and Sütterlin, 2014). Affect has been sub-categorised: Affect Heuristic I, where individuals are influenced by new information to interpret; and Affect Heuristic II, where individuals are influenced by their previous affect of a topic/technology prior to new information being received. Its influence may be reduced if individuals are required to justify their decisions to others (Siegrist and Sütterlin, 2014).

While there is substantial literature addressing cognitive biases, there is much less work concerning motivational biases. Both sources are hard to eliminate but Montibeller and von Winterfeldt (2015) suggest techniques to address potential sources of bias in both risk and decision analysis representations. However, their discussion only appears to consider steady-state situations where resource and time pressure are not present, so the techniques recommended are not necessarily compatible with aircraft maintenance scenarios. Mellors and Locke

[33]The conjunction fallacy occurs when there is a tendency to subset likelihood, believing specific conditions are more probable than a single general one; for example believing that the subset "men who are smokers" are more probable to have lung cancer than the set "men".

(2007) provided general suggestions for minimizing human mistakes within the decision process:

- *Combining Frames*[34] from multiple descriptions into a single frame may help to prevent preference reversals. More robustness (in safety applications in particular) may be achieved by framing the situation more toward the 'gain' frame, than the 'loss' frame, which encourages more risk averse behaviour.

- Describing probabilistic values in a *Frequency Format* (for example as "1 in 1000 chance") is clearer than a Bayesian inference format and can help reduce the chances of conjunction errors occurring.

- *Disclosing* conflicts of interest between personal and professional interests, leads to overestimations when subjects try to compensate for known bias or diclosure.

- Pre-decisional *Accountability* reduces: sensitivity to information order; susceptibility to conjunction errors; overconfidence; likelihood of being swayed by vivid (but irrelevant) or incomplete information; and improves willingness to revise opinions upon receiving unexpected feedback. However, decision-makers become more sensitive to status quo effects, are more likely to prefer the 'middle option' and compromise, and may be influenced by non-diagnostic information.

- *Incentives* have mixed effects. Although incentives can improve performance, in complicated tasks incentives can be detrimental to the overall situation. Incentivisation improves recall of items, easy problem solving and performance in clerical tasks, and reduces the effects of anchoring and adjustment. No rationality violations were made to disappear as a result of incentives.

[34] Framing describes the manner in which a statement is presented, for example the phrase "99% chance of surviving an average car journey in the UK" is positively framed (the gain frame); reframing the same situation negatively (the loss frame) would read "1% chance of dying on an average car journey in the UK". Decisions have been found to be influenced by framing (Bazerman and Moore, 2009).

– *Calibration Training* to improve accuracy can reduce overconfidence in probability estimates. Professional groups, such as weather forecasters and horse racing handicappers, have been found to be well calibrated because they receive timely feedback on prediction accuracy. Calibration is suggested to be improved by using debiasing techniques, scoring and reward systems, and systematic evaluation and feedback (Cooke, 1991, p26).

There is variability in descriptive decision performance between naive and applied situations. Engineers were found to have a strong tendency to avoid ambiguity when compared with a non-engineering cadre (Brown and Utley, 2019), although this study used a regionally-limited demographic with a test scenario that was not representative of a real-world situation. Using an engineering undergraduate cohort, Vermillion et al. (2015) found that engineers are likely to be risk averse with positively framed situations, but risk-neutral when situations were described in a negatively framed manner, which is contrary to findings by Kahneman and Tversky (1979). Another experiment using engineering student participants found their risk decisions to be based more strongly on likelihood than severity, to have underestimated failure probabilities, and their decisions were biased (anchoring and adjustment effects) by system failure and component reliability data (Gernand, 2018). These conclusions are restricted to untrained (undergraduate) engineers and may not adequately represent trained and practicing engineers. In another professional domain, Richie and Josephson (2018) used fictional but realistic real-world simulations to conclude that medical trainees were influenced by bias in domain-specific scenarios, suggesting that clinical experience reduces bias and heuristic effects. However, they also identified that outside of domain-specific scenarios, qualified and higher trained participants in their study were more prone to bias. In these papers, participants were tested with data that had certainty implied, which does not represent real-world decision problems, but nevertheless helps understand how humans (and engineers in particular) may react when presented with information.

Despite numerous documented cases of applied engineering judgement being bias, there remains no comprehensive systematic method for handling judgement

(Baybutt, 2018). There may be benefits to decision-makers recognising the existence of bias in assessments and engineering decisions could be improved if the influence of bias could be controlled. Unfortunately, Baybutt stops short of offering insight into the effectiveness of debiasing techniques in engineering hazard and risk analysis.

2.3.5 Prescriptive decision analysis

In real-world situations that face dynamic and changing decision environments, normative decision approaches methods are too cumbersome, time consuming, overly reliant on significant information being available and are generally not followed by decision actors (Mosier (2008), Crossley et al. (2017)). By identifying tools and methods to support decision-makers in dynamic real-world decision environments, prescriptive decision analysis lies somewhere between normative and descriptive theories. The prescriptive paradigm owes it's roots to Herbert Simon (Simon, 1955) who noticed that optimization is rarely possible in the real-world, and coined the term *bounded rationality* to highlight the manner in which humans make decisions when optimization is out of reach (Marewski and Gigerenzer, 2012). This was later extended to describe *ecological rationality*, which defines a form of rationality limited to the specific environment and goals of the decision-maker (Todd and Gigerenzer, 2012). The success of heuristics is dependent upon the "degree they are ecologically rational, that is, adapted to the structure of the information in the environment in which they are used to make decisions"(Gigerenzer et al. (1999) cited in Crossley et al. (2017, p515)).

Naturalistic decision-making (NDM) developed from a desire to study how experts were able to make effective decisions under complex and time constrained conditions, being able to "quickly identify a subset of cues that is critical to accurate diagnosis of a situation and the subsequent decision and action" (Mosier (2008, p41),Brown (2015)). It provides an alternative lens to laboratory-based descriptive theories, and has become a mainstream applied research paradigm, contributing to the development of applied cognitive research methods and

benefiting human factors and ergonomics applications (Klein, 2008). This highly relevant field offers potential insight into the specific decision-making context that is complementary to aircraft maintenance structural fault decision-making. One model, Recognition-Primed Decision Making is a hybrid blend of Type 1 and Type 2 (see Table 2.1) cognition and describes that individuals use their experience by a repertoire of patterns to diagnose a situation by the most relevant cues and causal factors, being able to make rapid decisions (Klein, 2008). A similar model is the Recognition/Macrocognition model (Cohen et al., 1996), which posits that decision-makers use generic knowledge about particular scenarios to integrate new information, creating connections between causal and intentional relations between events to build narrative story structures. Both of these attend to intrinsic constructions of situations based on experience, and were developed by observations of experts in the natural decision-making environments.

Other developments of NDM include Cognitive Task Analysis methods[35] and Macrocognition[36]. Like Polacsek et al. (2018) applied earlier (Section 2.3.3), the use of argumentation has also been found as a naturalistic decision strategy for industrial dispute resolution (Ramiah and Banks, 2015). This commonality indicates the usefulness of argument for domains that require justifiability of decision-making, such as those found in aviation maintenance.

Fast and Frugal Heuristics (FFH) perform the same task as NDM, namely to overcome limitations faced in complex and time constrained decision situations. However, FFH identifies ecologically rational heuristical procedures that humans apply to give quick and efficient decision support. They are commonly structured around three rules for identifying relevant information, knowing when to stop searching for more information, and acting upon the information found. These are described in general terms:

Search Rule in what direction does the search for information extend in a defined search space.

[35]Cognitive Task Analysis describes the methods available to elicit understanding of the cognitive processes in a task being studied.

[36]Macrocognition is the study of cognitive adaptations to complexity (Klein, 2008).

Stopping Rule defines when the information search is stopped.

Decision Rule specifies how the final decision is made.

Such models are aimed at *satisficing* by obtaining solutions that are sufficient to satisfy the decision-maker's goals, as opposed to *maximising* or *optimising* in some normative models. Such decision strategies are useful in practical circumstances because practitioners are more likely to use them (Hafenbrädl et al., 2016), so an 'adaptive toolbox' is the widely suggested way forward for operationalizing fast and frugal heuristics (Hafenbrädl et al. (2016), Marewski and Gigerenzer (2012), Gigerenzer (2014, p30)). Some heuristics applied by humans are 'rules of thumb' that are naturally developed, such as the Gaze Heuristic which animals use to catch prey by holding a constant angle between their line of motion and that of the target (Gigerenzer, 2014, p30). Others use rapidly employable methods to help prioritise information and make a selection, such as the 'Take The Best' heuristic, which relies on only the best cues (in the decision-maker's opinion) and ignores the rest. Hafenbrädl et al. (2016) discuss two heuristic methods ('Take The Best' and 'Tallying') that out-performed a complex method (Multiple Regression) in prediction (under situations of uncertainty), although the complex method performed better than both heuristic methods when data became known. There are numerous tools available, which Hafenbrädl et al. briefly describe, although there remains weak understanding of when heuristic strategies should be selected. In particular, there is a requirement to know the validity of circumstantial cues for predicting an outcome *a priori* - the uniqueness of structural damage means that such objective information is not expected to be available to aircraft maintenance decision-makers for every structural configuration, material, and applied loading nature possible.

2.3.6 Decision analysis summary

Decision analysis spans two research epistemological paradigms of objectivity and subjectivity, but both normative and descriptive fields offer valuable insight to help provide pragmatic, applied solutions. Prescriptive decision analysis fills

the space between the two and substantial work has been completed in this regard. However, no evidence is found of prescriptive decision analysis work that had been applied to structural integrity issues in aircraft maintenance. Infact, only very recent papers have offered engineering-specific descriptive decision insight, that professionals act in a different manner to those prescribed by Prospect Theory. Seeking an ecologically rational solution through heuristical methods offers promising ways forward for assessments of aircraft structural integrity faults in operational conditions. Utilising research from descriptive work on debiasing techniques and cognitive tendencies of engineers will also support development of a decision aid, and will contribute to the growing research effort to integrate Type 1 and Type 2 research to provide usable decision tools. Murphy and Pate-Cornell (1996) SAM model is one such effort in this area, combining both bounded-rational and rational models of action (Khan et al., 2015). However, the model appears only applicable at levels where control over management actions can be retained, so further work is required to demonstrate if such prescriptive methods are applicable in operational engineering decision situations.

2.4 Literature review findings

A review of the literature for relevant and related work was framed by a set of basic requirements outlined by the research context, aim and objectives. The review considered risk and decision literature as discrete sections.

Risk analysis is extensively applied in engineering applications owing to it's ease of application, clear criteria measurement and ability to be modelled analytically. However, in the case of calculating the risk from aircraft material damage, this is a difficult, intensive task and the methods identified are not suitable for operational maintenance situations under time, resource and information constraints. Academic approaches to generating accurate risk analyses of crack propagation or corrosion deterioration effect are laboratory-focussed, dependent upon knowing material variables and upon building an appropriate requisite model, requiring an expert understanding of fracture mechanics analytical methods.

This is an unrealistic expectation for aircraft maintenance practitioners, who are trained engineering managers but not structural experts. It is expected that in an operational situation demanding a maintenance decision about new and unexpected aircraft structural damage, judgement is applied. One of the clear findings from the risk analysis literature was that the context of the risk analyst and decision-maker was rarely taken into account in the development of risk analysis methods for structural integrity. No secondary data was identified that confirmed decision-maker behaviour regarding structural integrity risks. The lack of a coherent approach to the treatment of uncertainty (in particular deep uncertainty) means that a justifiable approach to handling uncertain variables in structural calculations is unavailable. This is further exacerbated by the lack of validation against real-world situations in the applied literature reviewed.

In a similar perspective to validation, few literature sources justified the transparency and auditability of their methods. Bayesian networks were favoured for their hybrid approach of encoding both qualitative representation and quantitative measures appropriate to the situation. However in analytical applications, subjective choices were rarely made clear, such as the selection of probabilistic distributions for variables or the critical crack length criteria for material failure. The proliferation of simple risk analysis and management techniques, such as risk matrices, demonstrates the desire for easy-to-use methods but has meant that problems with those approaches have become widespread and systemic.

Decision analysis paradigms offer alternative ontological perspectives regarding aircraft maintenance structural integrity faults. Existing industrial expectations to measure risks against a quantifiable threshold restrict the solutions available given the time, resource and information constraints felt in an operational environment. Normative approaches are useful for providing a structure for decision-makers to use, although are not necessarily followed in real situations. Prescriptive approaches may be more suited to the decision context and time horizon available through structured accounting for biases and heuristic decision shortcuts. Understanding the influences upon a decision-maker identified by descriptive decision analysis may assist decision-makers in ensuring their decision

remains as ecologically rational as necessary for the given situation, especially knowing that engineers have traits that differ from naive participants.

The literature review has highlighted a lack of practitioner-facing risk decision-making approaches that are cognizant of situational factors. No secondary data describing the aircraft structural integrity decision-making context was identified, confirming the need for Objective 1. A gap in the literature also exists for a naturalistic approach that can help practitioner decision-makers deliberate and apply their judgement about complex aircraft structural damage, which acknowledges the constraints of an operational maintenance environment. This supports the intent of Objective 2. Ensuring that such a method is auditable preserves public accountability, but requires subjective judgements to be disclosed. The decision must meet the goals and objectives specific to each decision-making circumstance and achieve a decision basis that can satisfy regulatory scrutiny. Addressing both objectives fulfils the overall research aim to better support aviation engineering decision-makers, and highlights the need to diversify the literature with naturalistic and prescriptive decision-making research.

2.5 Chapter summary

A literature review of risk and decision analysis material, relevant to structural integrity risk decision-making, has been carried-out. Scrutiny of literature was framed by a set of contextual requirements that enabled evaluation of literature relevant to the research aim. Although the domains of risk and decision analysis are both extensive, no research was found that addressed improvements to aircraft structural integrity risk decision-making. Methods identified that addressed structural integrity specifically were complex, time-consuming and required specific expertise to build, validate and execute models. Decision approaches that concerned naturalistic and constrained situations provided descriptive insights but did not offer a solution to supporting practitioners compelled to make decisions about structural faults without access to formal analysis. No recent data assessing maintenance risk management practices was identified either. The observations

from the literature review support both research objectives.

Chapter 3

Initial data collection methodology

As a qualified and experienced maintenance manager, the difficulty experienced when constructing a justifiable risk assessment about military aircraft structural risks was the author's motivation behind the research project. Organisational aspirations to manage threats to human life and operational capability are well intentioned until a situation emerges where the assessment method is incompatible with the problem context. This is the case for aircraft structural faults found during maintenance, where punctual decisions are required to sustain military aircraft availability, but a meaningful and transparent assessment of risk is rarely possible without substantial down-time. The application of unsupported, subjective judgement, conditional on individual training, experience and knowledge, is (in the author's experience) the routine fall-back. While whole structural system failures are (thankfully) rare, the lack of transparency or guidance in the application of judgement is one accident away from becoming an urgent need. This chapter describes the methodology used for primary and secondary data collection, in order to rigorously identify structural integrity risk management practices applied under operational constraints.

3.0.1 Initial research question

The author's experience of risk assessing structural faults under operational conditions, informed a broad initial research question, which was necessary to

focus early work during the research project:

> *What is the most effective strategy for enabling decision-makers to assess current and future structural integrity risks?*

Both Phases 1 (Chapter 2) and 2 (Chapters 3 and 4) were framed by this question and it enabled a broad knowledge-base to be established through the literature review and initial data capture.

3.1 Initial data capture and analysis (Phase 2)

Phase 2 covered primary and secondary data collection and analysis (Chapters 3 and 4) in order to inductively assess naturalistic practice and inform the development of a decision support framework in Phase 3 (Chapter 5). This was achieved through surveys of structural integrity practitioners with experience of risk decision-making, and by identifying relevant case studies to support survey outputs.

3.1.1 Data collection purpose

Phase 2 adds depth and academic rigour to the author's experiences of how structural integrity risk analysis was conducted in practice. Primary data collection from real-world risk decision-makers also supports the call for more empirical evidence of real-world activities (Rae et al., 2020). In Phase 2 of the research, both primary and secondary data were collected for analysis. Primary data consisted of survey and interview responses and a case study involving a UK MOD aircraft. Secondary data consisted of two further case studies, already in the public domain.

The survey was planned to capture a detailed description of how practitioners treat structural integrity risks when constrained by time, information and resource, as well as the broader organisational risk management of structural integrity. This was intended to highlight conditions where existing risk practices were lacking attention and support.

Case studies provided retrospective evidence from risk decision-making situations under real-world operational constraints. Since survey data is subjective, self-reflective and can be considered as unreliable (Foddy, 1993, p1-11), case studies provided complementary evidence of industrial practice to the survey.

3.1.2 Ethical approvals

Prior to Phase 2 data collection, ethical approvals were required from both Cranfield University and from the UK Ministry of Defence (MOD). Ethical approval from Cranfield University proceeded according the University's CURES protocol, with approval being granted[1]. The process required for the UK MOD was longer than anticipated and led to the planned schedule of the research being extended during Phase 2 (Appendix A). The MOD Research and Ethics Committee (MODREC) reviewed both methodological and ethical aspects of the planned Phase 2 data collection, and necessitated an in-person scrutiny by a panel including professional and academic representatives. Approval was granted following further clarifications and adjustments, including a second showing at the scrutiny panel. The process from initial application to approval took 7 months[2]. Bespoke consent forms for both approvals systems were included within the primary survey data collection.

3.2 Survey data collection design

3.2.1 Data sources

Although general risk attitudes are well studied in the literature (Short and Sikora (2016), Chionis and Karanikas (2018)), no recent representative secondary data were found that could inform a study of risk handling by structural integrity practitioners in real-world aircraft maintenance operational scenarios. Any

[1] CURES project ID "6852".

[2] The MODREC application number pertaining to the study is 950/MODREC/19, which relates to all Phases of the research project.

secondary data would have needed to be recent in order to be representative of current industrial practice. It was therefore necessary to gather original primary data from professionals who make structural integrity risk decisions through their employment, in order to gain genuine naturalistic perspectives of risk decision-making as it is applied in real-world conditions. Additionally, targeting practitioners within their industry means that the participants are (or have been previously) trusted to make appropriate risk decisions within their organisations, and therefore are (or have previously been) accountable or responsible. In comparison to many general studies into human decision-making, a more focused naturalistic study can provide a rich data source to study real-world decision settings that include information ambiguity, tangible consequences (to society or individuals), and meaningful measures that have familiarity with practitioners. Although laboratory settings are intended to remove such variances in order to isolate variables of interest, real-world events that influence decision-maker attention should not be "dismissed as just experimental nuisance factors" (Payne et al., 1993, p173). Surveys of specific cohorts are not unusual in the literature (Richie and Josephson (2018), Brown and Utley (2019)). However, unlike these studies that aimed to evaluate a specific effect, this data collection exercise was aimed at inductively identifying the attitudes and practices toward structural integrity risks and verifying the direction of the study.

The survey sought perspectives from a variety of organisational functions within the equipment life cycle, namely: operators/end users, manufacturers/designers and regulators/policy makers. This diversified responses beyond just a "sharp end" viewpoint. For instance manufacturers manage structural failure risk by making assumptions about future usage during design. This is effective mitigation until through-life usage design assumptions become invalid (Yan and Liu (2011), Tavares and de Castro (2017), Yates and Stone (1992, p304), Section 1.2.2).

3.2.1.1 Selection criteria

Specific inclusion and exclusion criteria were established to ensure that participants had experience of real-world engineering structural integrity

decision-making. Questions were intended to elicit opinions and understanding of organisational policies and individual practices, hence participant responses that could not contribute to this objective would dilute the conclusions. Criteria were included in invitational correspondence.

- *Inclusion criteria* Participants who self-identified as engineering managers, trained and experienced in the application of decision-making and management of structural risks as either an analyst, a decision-maker, an advisor or a commentator. Any participants whose suitability was questionable could be identified through preliminary filter questions within the survey, and failing that through their use of language and detail given in responses to qualitative questions within the survey. No responses were required to be discounted.
- *Exclusion criteria* Participants were not invited to participate if they were untrained engineering managers, i.e. had not reached a company-recognised level of competency, or if they were trained engineering managers but had had no exposure to structural risks.

3.2.1.2 Recruitment method

Targeted participants were approached by email via focal points within the UK MOD, with a request for onward distribution in a hub and spoke-type approach, see Appendix J. An opportune conference internal to the UK MOD Royal Air Force Engineering Branch enabled a short verbal and poster presentation to be given (Appendix B). Participants outside of the UK MOD were invited to participate through networking and engagement events, and by requesting distribution from professional engineering insitutions[3]. Social media channels were not exploited since it was felt that a targeted recruitment campaign necessitated a personal approach, as far as possible. Furthermore, although social media channels may improve total visibility of the invitation to participate, it may have also had the

[3]Professional institutions approached were: The Institute of Engineering and Technology; The Institute of Mechanical Engineers; and The Royal Aeronautical Society.

negative impact of increasing unsuitable participant attempts to complete the survey.

Sample size and incentivisation In order to provide the opportunity to make generalisable conclusions from the data and provide meaningful confidence bounds using standard statistical tests and assumptions regarding normality, at least 30 participants were required. Participants were incentivised to complete the survey and provide contact details for follow-up engagement in the research project in exchange for their entry into a draw to win a retail voucher. It was perceived that participants' willingness to complete the survey would indicate appetite to support later research participation. Of the 51 participants that commenced the survey, 32[4] fully completed it, and of those 21 provided their contact details. The conduct of the draw was scrutinised for completeness and carried-out by an independent university staff member in the presence of a witness, with the winner being notified by email of their prize.

3.2.2 Survey format and design

A structured questionnaire provided a specific point-in-time data set detailing real-world practices, perspectives and beliefs. Interviews with 7 practitioners from aviation and rail domains were undertaken to pilot the question set for ambiguity, interpretive differences or other errors in data collection. Feedback from the pilot testing led to adjustments in question wording and the order of questioning.

3.2.2.1 Pilot testing survey design

An initial collection of 50 questions (not including sub-questions) followed from the initial research question in Section 3.0.1 and were grouped to enable logical flow of questioning as follows:

– Context and Bracketing: for grouping participants by experience, domain and industry.

[4]Counted from Question 35, which was the last closed question of the survey.

- Stakeholder involvement and structural integrity management, communication and policy.
- Belief in structural integrity risk management: identify how far participants believe in the risk management system used in their organisation for handling structural integrity risks.
- Substantiating structural integrity: identifying participant perspectives on how structural integrity risks are substantiated in operational situations, including risk parameters and quantities that assist in the appraisal of risk (uncertainty, assumptions, conservatism, use of judgement).
- Supporting the decision-maker: identify mechanisms within the participant's organisation that provide support to the decision-maker's structural integrity risk assessment.
- Respondent risk attitude.

Quantitative questions were worded to enable measurement by Likert scale or categorical comparison between marginal groups. Data collected were ordinal or nominal. Qualitative questions were set to enable rich descriptive responses, such as describing the method for conducting risk analysis, the information that participants required for making a decision on a risky prospect, or to provide context to a quantitative-style response to the question immediately preceding.

Pilot testing indicated that improvements were necessary to reduce the survey length, to remove questions that were supplementary to the research focus, and to re-order the questions to improve survey flow. Questions that were removed regarded the topics of risk management framework maturity and participant risk attitude.

3.2.2.2 Main survey design

Following the pilot, refinement of the survey sections became:

- Context and Bracketing.

- Analysis aspects of in-service structural integrity risk.
- Decision-making specifics of structural integrity risk.
- Management of structural integrity risks arising in-service and the handling of risk analysis conducted.

Appendix D provides the full survey text and answer options presented to participants. The resulting question set contained 37 leading questions[5]. The estimated completion time communicated to participants was 30 minutes, which was an underestimation as the median completion time for participants was 40 minutes. The long duration of the questionnaire was considered a risk to participant completion rate. These concerns were partly realised, with 48 participants completing Question 5 (last question of survey Section 1), dropping to 32 participants completing Question 35 (the last closed question of the survey), a 66% completion rate.

3.2.2.3 Marginal group selection

Predictor variables were chosen so as to be broad and preserve respondent anonymity: industry/domain; organisation function; and participant role. These are related to the levels on which human factors act: group, organisational, and individual levels respectively (Hillson, 2017, p16).

- *Industry.* Aviation, civil, automotive, nuclear, rail, other (specify)
- *Organisation functional category.* Operator / end user, regulator / policy-maker, designer / equipment manufacturer[6].
- *Individual role.* Open ended answer, to contextualise answers if deemed necessary.
- *Organisational policy.* Whether organisation risk management policies specifically consider structural integrity risk.

[5] Including sub-questions the total number of questions within the questionnaire was 84, although exposure to all 84 was not possible owing to the question options set.

[6] Tavares and de Castro (2017, p1515) state that “airworthiness involves the interplay of authorities, manufacturers, operators and maintenance providers”.

– *Individual role related to risk.* Risk analyst [7], risk decision-maker[8], commentator[9], advisor[10].

3.2.2.4 Data collection platform

Pilot test survey Pilot test data were collected by interviewing pilot participants either in-person, by telephone or by video conferencing, using a semi-structured interview format. Interviews were audio recorded with verbally granted permission, in order to reflect on the conversations for potential opportunities not identified in the draft question set. Although a semi-structured format resulted in some questions being overlooked during interview owing to limitations on time afforded by participants, it enabled the author flexibility to explore unexpected areas of interest with the participant, which was considered a beneficial compromise during pilot testing. As the first engagement with real-world practitioners, this was helpful for identifying potential problems or opportunities ahead of the primary Phase 2 data collection. The interviews also served as an initial proving that the research issue was genuine and not just the belief of the author. Transcripts from three of the pilot interviews are contained in Appendix C, which demonstrate the interview conduct and types of responses provided.

Main survey The survey was designed to be compatible with the Qualtrics survey platform (Qualtrics, 2020), which provided a convenient online portal for participants to access questions through any internet-enabled platform (desktop, laptop, mobile or tablet). No participants provided any feedback pertaining to survey access difficulties. Appendix E Figures E.1 to E.3 display examples of how the survey appeared visually in Qualtrics to participants.

[7] Defined as an individual who produces risk assessments based upon observations, beliefs and information available at the time of the analysis.

[8] Defined as an individual who decides upon the course of action for a given risk. They may make consideration given other factors outside of the particular risk envelope, e.g. operational, financial, commercial or other business constraints.

[9] Defined as an individual who may take a view on a given risk but has no particular responsibilities within the risk management framework.

[10] Defined as an individual who may take a view and provide advice to a risk analyst or risk decision-maker, but is not responsible for the decisions made within the risk management framework.

3.2.3 Survey anticipated outputs

To help meet Objective 1, the survey was required to:

- Rigorously confirm the systemic shortfalls in risk analysis practice suspected by the author, regarding an inability to quantify structural integrity risks given contextual constraints.
- Inductively analyse naturalistic industrial structural risk practice under situational constraints.

The specific outputs anticipated were intended to help achieve Objective 2 by guiding the development of a novel decision support for risk decision-makers, requiring:

- Identification of typical understanding of structural integrity risk management within stakeholder communities.
- Survey particular processes or techniques in use for managing structural integrity risk.
- Highlight fallibilities in user-applied risk processes.
- Identification of specific structural risk management requirements under operational conditions.

The handling of responses required tailored considerations: answers providing evidence of artefact usage (for example particular risk analysis methods) provided plain data; responses requiring a participant's subjectivity were treated with caution owing to the potential for respondent bias. Although difficult to make general inductive conclusions from subjective data, individual perspectives were expected to be sufficient to justify any findings when there was consistency identified between respondents, and when combined with case study findings. Follow-up interviews were retained as an option if deemed necessary to garner greater detail, although this requirement was not exercised.

The integrity of the survey responses was dependent upon how far participants believed that their responses would be treated with confidentiality and anonymity.

Safeguards within the survey reduced the likelihood of these effects, such as by not requiring specific detailed information, limiting contact information to voluntary only, and being clear with respondents in invitations about how their contact details and responses would be used and protected.

Independent review limitation Owing to the nature of the PhD award, no second rater was employed on the qualitative data analysis. Second raters are useful in thematic analysis to provide alternate perspectives on themes identified (see for example Provan et al. (2019) and Morgan et al. (2016)) but this cross-checking was not available. Consequently, this is a limitation of the findings from Phase 2.

3.2.4 Analysis methods

Some commonality between a number of the pilot survey qualitative questions and the main survey enabled qualitative responses from the pilot to be included in analysis. However, pilot survey quantitative responses were not included owing to dissimilarity in question order, wording and options presented to pilot participants. This is because the order of questions has bearing on the consistency of answers given, with earlier questions influencing how subsequent questions are responded to (Foddy, 1993, p60).

3.2.4.1 Survey qualitative data

Every survey question that invited a descriptive answer was reviewed using thematic analysis techniques to draw together common themes and patterns from the information provided by respondents. While drawing together common themes would reduce the influence of respondent bias, the connections between responses were made by me. Although my experience was essential to appropriately connect related context-specific fields, the subjectivity of the generated themes is open to challenge, compounded by the lack of a second rater.

Thematic analysis follows a loose and flexible structure (Saunders et al., 2016,

p579), so the discrete steps taken in the Phase 2 results analysis are outlined as follows:

- Organise collected survey data into tabulated form.
- Extract qualitative questions and segment into individual questions, alongside relevant confounding information.
- Review answers, extracting themes, cues, common phrases, behaviours, attitudes, beliefs, actions, technique usage. Codify theme headers and label source data for traceability.
- For each question, group themes into higher order categories of data if appropriate.
- Summarise themes identified within categories and highlight the common findings for each question.
- Identify the qualities of each question that can be taken forward to support methodology development.

Unlike template (Saunders et al., 2016, p587) or content (Saunders et al., 2016, p609) analysis, which rely upon a pre-existing structure to organise qualitative data, no common structure exists with which to support the analysis, so codes were generated as the data were analysed. The survey contained 36 qualitatively-focussed questions that required thematic analysis. This generated a significant analysis burden to organise, codify and extract the meaningful information to proceed with. Thematic maps were the main output from qualitative questions, which linked themes together into categories relevant to the subject of the question.

Codification was carried out for each participant's answer using hand-written notes to capture the summary descriptors of the discrete process, function or grouping deduced from the data item. Relevant codes were framed by the question topic, for example Question 20 required participants to "Describe the decision-making process regarding in-service structural integrity risks". In this case

relevant data was only codified if considered related to a decision-making process, producing codes such as “calculate/determine risk”, “identify resources necessary for mitigation”, and “identify other people for consensus” (see Figure 4.18). A single data entry could attract multiple codes, if applicable. The hand-written notes were then arranged according to linking themes or ideas that connected related data entries from other participants. Photographs taken during typical coding and sorting are shown in Appendix G, Figures G.1 and G.2.

3.2.4.2 Survey quantitative data

The survey questions that provide quantitative data were either nominal or ordinal Likert scales of preference. Likert scales were converted into numerical equivalents of equal weight, in order to enable numerical manipulation. Results were primarily descriptive in nature, with confidence intervals provided to enable generalisation of the results to engineering structural integrity risk practitioners. Although the questions were not designed explicitly to compare two independent sample groups, parametric and non-parametric tests of statistical significance have been carried-out to check for any large effects between predictor variables. The sample sizes were not sufficiently large enough to identify anything other than large effects owing to a lack of statistical power given the number of participants. For Likert data comparisons, the Mann-Whitney U Test was used to provide statistical comparison of two marginal data groups, for example comparing the “operator/end user” group against non-“operator/end user” group (an aggregated group made up of “regulator/policy-maker” and “designer/manufacturer” groups). Where there were binary categorical data, a simple test of sample proportion differences was applied using Fisher’s exact method.

Descriptive results analysis was carried-out using Microsoft Excel (Microsoft Corporation, 2020). Statistical difference testing was carried-out using JASP (JASP Team, 2019) and Minitab (Minitab, 2020), which required the extraction of quantitative data from the complete raw data in .csv format.

3.2.4.3 Data validity

The limitations of independent study review (Section 3.2.3) and the nature of the research data being of naturalistic inquiry invite perspective on data validity. Lincoln and Guba (1985, p289-301) argue that a naturalistic paradigm requires an equivalent view of conventional validation through “credibility”, “transferability”, “dependability” and “confirmability”. Alongside the peer checking (through the academic award process), narrowing of the research question following Phases 1 and 2, case study triangulation of specific conclusions (Section 4.3), and “member checks”[11], the trustworthiness of the data generated by the survey is established. In particular, the member check was conducted by aggregating survey results to directly inform the framework design choices during the development of the novel decision support tool during Phase 3. Successful application of the proposed decision support solution evaluated during Phase 4 with representative participant testing gave respondents the opportunity to demonstrate utility, provide further feedback, and ultimately confirm the credibility of the conclusions drawn from Phase 2.

3.3 Case study data collection design

3.3.1 Data sources

Data collection through case studies were appropriate to provide a balanced perspective on industrial practice, given the survey data required participants to be self-reflective, and may contain motivational or cognitive biases in the responses given. Coherence between the findings would amplify the strength of conclusions.

With experience of the aviation maintenance structural integrity domain, the author exploited professional connections in order to identify relevant military aviation case studies initially, but expanded to include other transport domains.

[11] A member check is a process where “data, analytic categories, interpretations, and conclusions are tested with members of stakeholding groups from whom the data were originally collected” (Lincoln and Guba, 1985, p314).

The process of recruitment for the survey provided further opportunity to request consideration of suitable case studies from participants.

In order to provide contextual consistency, candidate case studies required criteria to be of relevance to the research concentration:

- References an *engineering structural integrity risk decision-making* situation.
- Is contextualised by *an established risk management system.*
- Is subject to *operational environmental constraints* upon decision-making.
- Details the involvement of a *human component* and their selection of a necessary intervention.

3.3.2 Case study data analysis

Case studies selected were analysed using a template analysis approach (Saunders et al., 2016, p587), with codes adapted from (Aven, 2018b). This recent paper provides usable descriptors for how a human actor may perceive risky situations, framed by a discussion of Type 1 and Type 2 thinking with regard to risk analysis. The codes (see Table 3.1) were applied to sections of text within the selected case studies to indicate where a particular code was applicable. No additional codes were added, but an adaption to the established taxonomy was made for [NBPA], which highlights the distinction between unsupported judgements and those based on professional analysis[12]. The analysis of the case studies was reported in Green et al. (2021) (Appendix F).

3.4 Chapter summary

The methodological approach for collecting primary and secondary data relevant to the research aim, and Objective 1 specifically, has been detailed. Survey data

[12]Original source reads “Professional judgements of how large the risk is, based on historical data and probabilistic analysis” and “Professional judgements of how large the risk is, by comparing with other activities” (Aven, 2018b, p238).

Code	Description	Reference
[LA]	Learned associations between ideas	(Aven, 2018b, p238)
[PERC]	Perception of risk as high, fearful and unacceptable	(Aven, 2018b, p238)
[RHI]	Judgements of the risk as high	(Aven, 2018b, p238)
[RTHI]	Judgements of the risk as too high	(Aven, 2018b, p238)
[UNHI]	Judgements of the uncertainties as high	(Aven, 2018b, p238)
[UNTHI]	Judgements of the uncertainties as too high	(Aven, 2018b, p238)
[NBPA]	When judgements of how large the risk is are not based on a professional assessment of how large the risk based on either: historical data, probabilistic analysis, or by comparing with other activities	(Aven, 2018b, p238)
[BGKN]	Use of unsupported thinking in professional judgements, such as risk aspects that may be hidden in K (background knowledge) like assumptions.	(Aven, 2018b, p239)

Table 3.1: Expressions used to codify the case studies by identifying unsupported decision judgements in risky situations.

was planned to be captured from non-naive practitioners, experienced in structural integrity risk decision-making. Qualitative survey responses were analysed using a thematic analysis approach. Quantitative survey responses in the form of Likert and categorical data were analysed descriptively or using standard statistical tests for comparing differences between groups of data. Secondary data from two representative case studies in the public domain, plus primary data from a third case involving a UK MOD aircraft, were collected to support the survey data, by corroborating any patterns or themes identified. The case studies selected met broad criteria, which were contextually aligned to the survey questions, and were framed by an initial research question. Case studies were analysed using a template analysis, with parent codes provided by Aven (2018b), which provided

appropriate expressions for individualistic judgements applied to risky situations.

Chapter 4

Survey and case study results and discussion

4.1 Practitioner survey results

The primary aim of the survey was to elicit practitioners' views on the use of judgement in the area of structural integrity risk and decision-making. Presentation of the results follows the same order that the survey was presented to the participants[1].

4.1.1 Survey section 1 - context

51 participants commenced the survey and 32 fully completed it.

Question 1, 2 and 5 - organisation type and participant's role within it The core demographics of the participants favoured aviation practitioners, operator/end user categories, and risk decision-maker roles (Figure 4.1). The domain representation indicates the participants who completed the survey, discounting partial completions. Owing to the nature of opportunity sampling (Section 3.2.1.2), the proportion of responses were similarly skewed towards aviation. Question

[1] Some questions have not been included in the presentation of results within the main body, but are instead included for completeness in Appendix H.

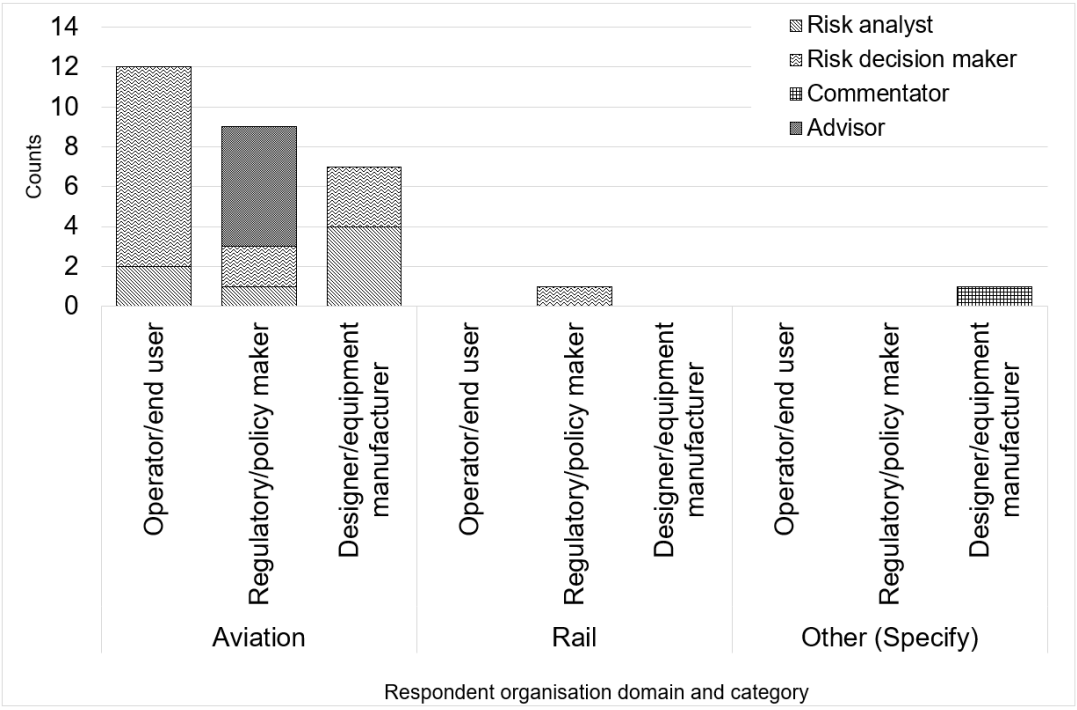


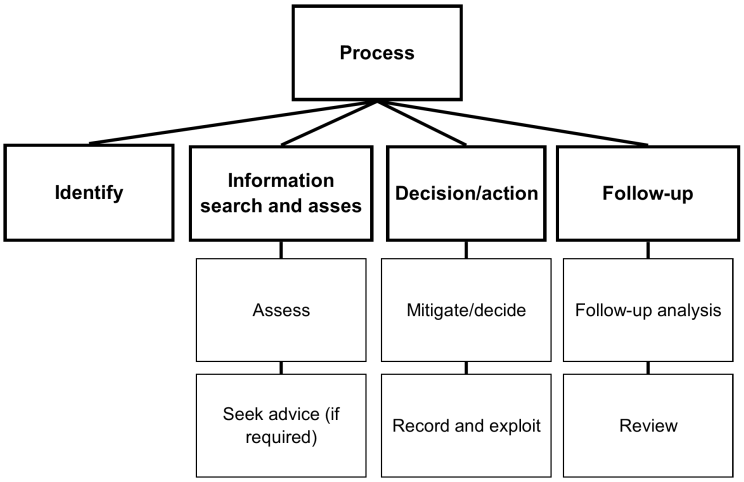
Figure 4.1: Practitioner survey demographics, illustrating participant role and categories. Data taken from survey Question 35, $n = 32$, which was the last closed question of the survey.

3[2] results have not been included in this results section as the responses were individualistic, specific to each respondent, and served to help frame understanding of participant responses given their description of their roles.

Question 4 - "In your experience, do organisational risk management policies specifically consider structural integrity risk?" "Yes" was the modal response (83.7% (41/49) $\pm$ 10.4% (95% confidence interval)), but with a 16.3% (8/49) ($\pm$ 10.4%, 95% confidence interval) "no" response, this indicates that a proportion of the population sampled does not believe that structural integrity risk is specifically considered within organisational risk management policy. Despite all regulator/policy maker responses indicating "yes", there were no significant differences between proportions (using Fisher's exact method).

Survey section 1 summary Participant demographic was overwhelmeingly representative of the aviation domain. Within the aviation domain, a majority

[2]Question 3 - "What is your role within the organisation?"



(a) Thematic map of responses relevant to processes.

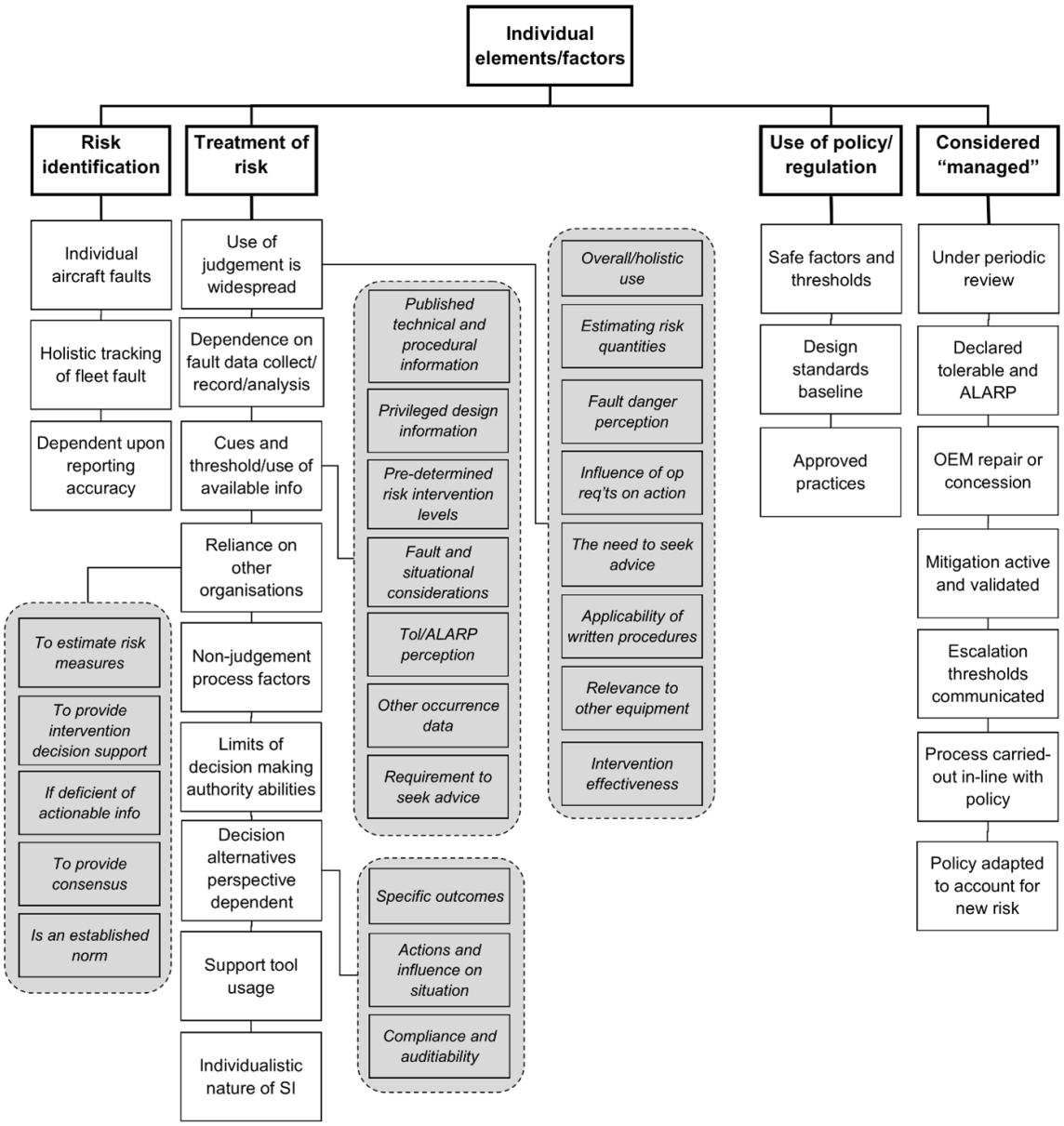
Figure 4.2: Survey Question 6 qualitative analysis.

of participants self-classified themselves as risk decision-makers from the operator/end user category. Excluding partial completions, 32 participants completed the survey, and a majority of participants felt that structural integrity risk was considered by organisational structural integrity risk management policies.

4.1.2 Survey section 2 - aspects of in-service structural integrity risk analysis

Question 6 -"'Describe the current process for conducting risk analysis for in-service structural integrity issues, from identification/ realisation of an issue to the point when the risk is considered managed." The free text answers were thematically analysed. Holistic themes relating to the *process* of conducting risk analysis were identified and are shown in Figure 4.2a. Discrete components that related to specific actions, activities or behaviours within the risk analysis process were also identified separately, shown in Figure 4.2b.

Question 7 - "In my organisation, structural integrity risk analysis is an activity that is valued." The modal Likert category was "agree" with 47.6% (20/42) $\pm$ 15.1% (95% confidence interval). Aggregating responses by whether they "agreed" to any extent with the statement, or did not agree (i.e. were "neutral"



(b) Thematic map of responses relevant to individual analysis elements and process factors.

Figure 4.2: Survey Question 6 qualitative analysis (cont.)

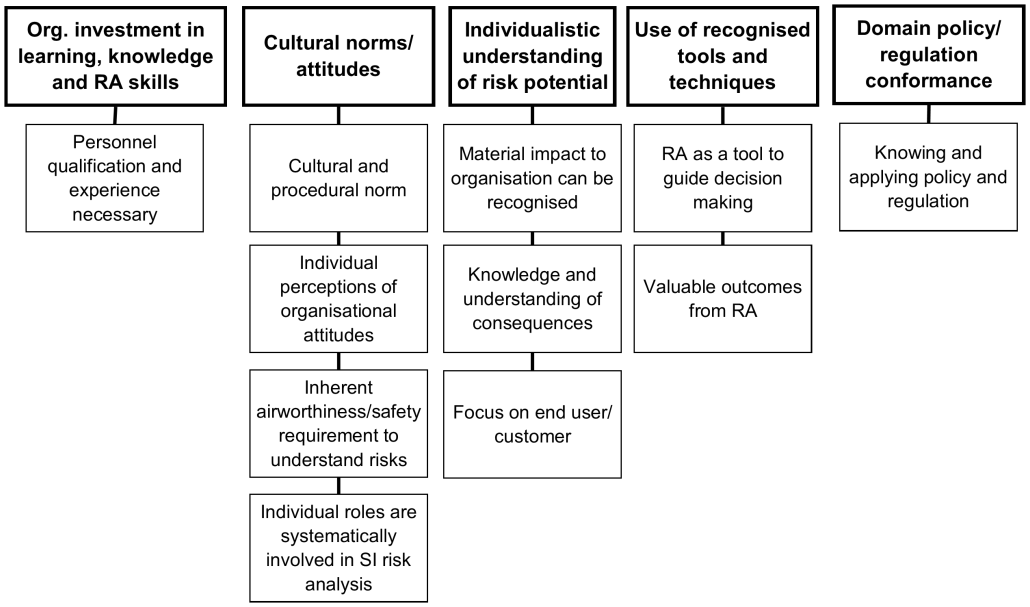


Figure 4.3: Survey Question 7a thematic map.

or “disagree” to any extent), 97.6% (41/42) of respondents agree (at least to some extent) with this statement.

Question 7a - “Why do you believe this to be the case?” Thematic analysis identified 5 primary themes surrounding the reasons why structural integrity risk is valued as an activity: institutional investment in learning, knowledge and skills in risk analysis; cultural norms and attitudes; individual understanding of risk potential; the use of recognised tools and techniques; domain policy and regulation demanding conformance. These and the sub-themes are shown in Figure 4.3.

Question 8 - “Within your organisation, are structural integrity risks that arise in-service described numerically or non-numerically / verbally / descriptively?” As displayed in Figure 4.4, ’both’ was the modal response (68.4% (26/38) $\pm$ 6.7% (95% confidence interval)), followed by ’verbally’ (29.0% (11/38) $\pm$ 6.5% (95% confidence interval)). Only 2.6% replied ’numerically only’. This pattern was consistent across category and role predictors.

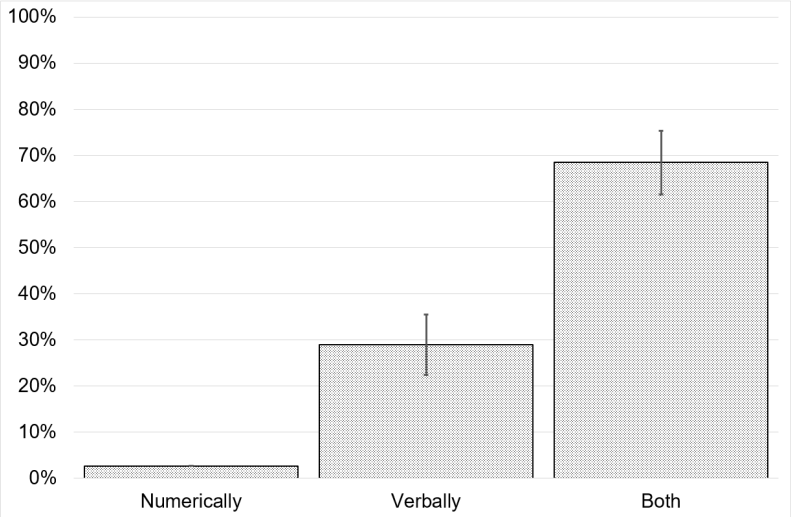


Figure 4.4: Relative frequencies of Question 8 responses - "Within your organisation, are structural integrity risks that arise in-service described numerically or non-numerically / verbally / descriptively?"

Question 8a - "How would you rate the accuracy of your organisation's quantitative risk analysis of structural integrity?" *(Only "numerically" or "both" responses from Question 8.)* The modal response was "Fair" (37.0% (10/27) $\pm$ 18.2% (95% confidence interval)), and 85% of all responses were "Fair" or higher. 4 respondents answered "poor" or "very poor", 3 of them were from the risk decision-maker role. Similarly, 3 of the 4 were from the operator/end user category. Designer/equipment manufacturer respondents gave higher accuracy ratings, and the mean difference was statistically significant at the 5% level (two-sided Mann-Whitney test, W = 99.0, p = 0.014).

Question 8b - "Why is structural integrity risk not quantified?" *(Only "verbally" responses from Question 8.)* Thematic analysis of answers identified the themes connecting the responses given: the nature of structural mechanics; organisational or cultural processes; time specific and situational factors; processing difficulties; and miscellaneous responses that were uncommon and didn't fit the other themes. The sub-themes are shown in Figure 4.5.

Question 8c - "In my opinion, a quantitative (numerical) value is necessary for the assessment and management of in-service structural

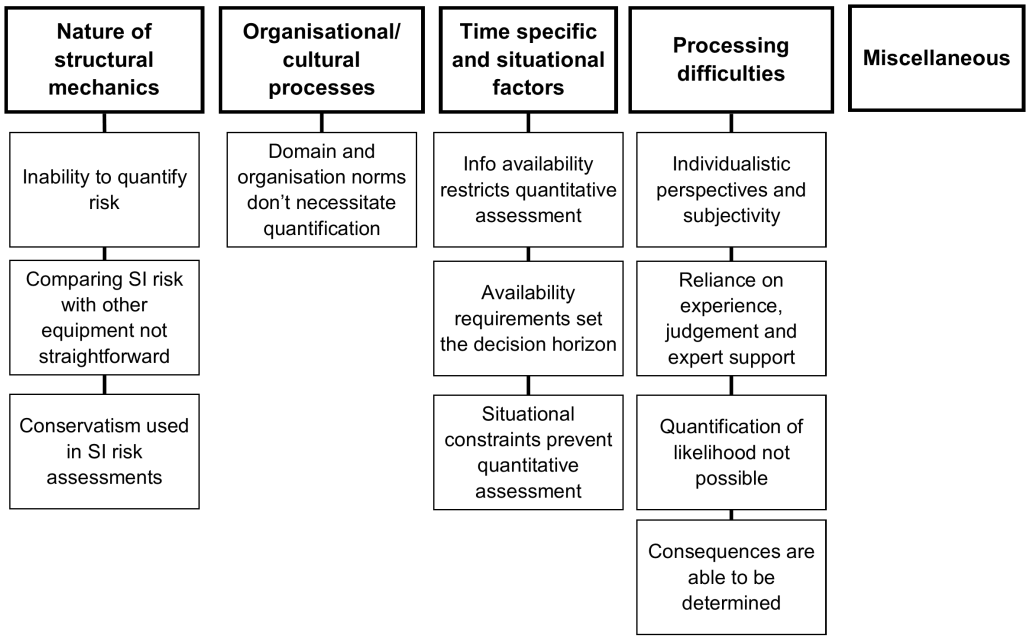


Figure 4.5: Survey Question 8b thematic map.

integrity risk?" The modal response was "agree" with 36.8% (14/38) $\pm$ 15.3% (95% confidence interval). Aggregating the results, 71% (27/38) $\pm$ 14.4% (95% confidence interval) agreed (to some extent), while 29% (11/38) $\pm$ 14.4% (95% confidence interval) of participants do not agree (to some extent). "Neutral or disagree" responses were seen from all respondent categories and roles (excluding commentator), and there were no significant differences identified.

Question 8d - "In my opinion, a qualitative (verbal/descriptive) interpretation is necessary for the assessment and management of in-service structural integrity risk?" The modal response was "agree" with 52.6% (20/38) $\pm$ 15.9% (95% confidence interval). Responses were grouped as for Question 8c. Only 5% (2/38) participants did not agree (both responses were "neither agree nor disagree") that a qualitative interpretation is necessary for handling structural integrity risk.

Question 8e - "Do you have any comments to support or expand upon these answers?" Comments provided a richer view on the answers given in Question 8 and it's follow-on questions. Thematic analysis connected themes

relating to practical properties of quantitative and qualitative methods, and how respondents select their approach. Thematic maps are shown in Figure 4.6. Notable responses from some operator/end users stated that quantitative outputs were not helpful, combined with a perception that quantitative risk values were useful for manufacturers - a perception not consistent with designer/equipment manufacturer responses to Questions 8c and 8d. Quantitative forms were stated to have drawbacks such as disguising background information and requiring confidence in the accuracy. For example Participant 4 (aviation, operator/end user, risk decision-maker) noted:

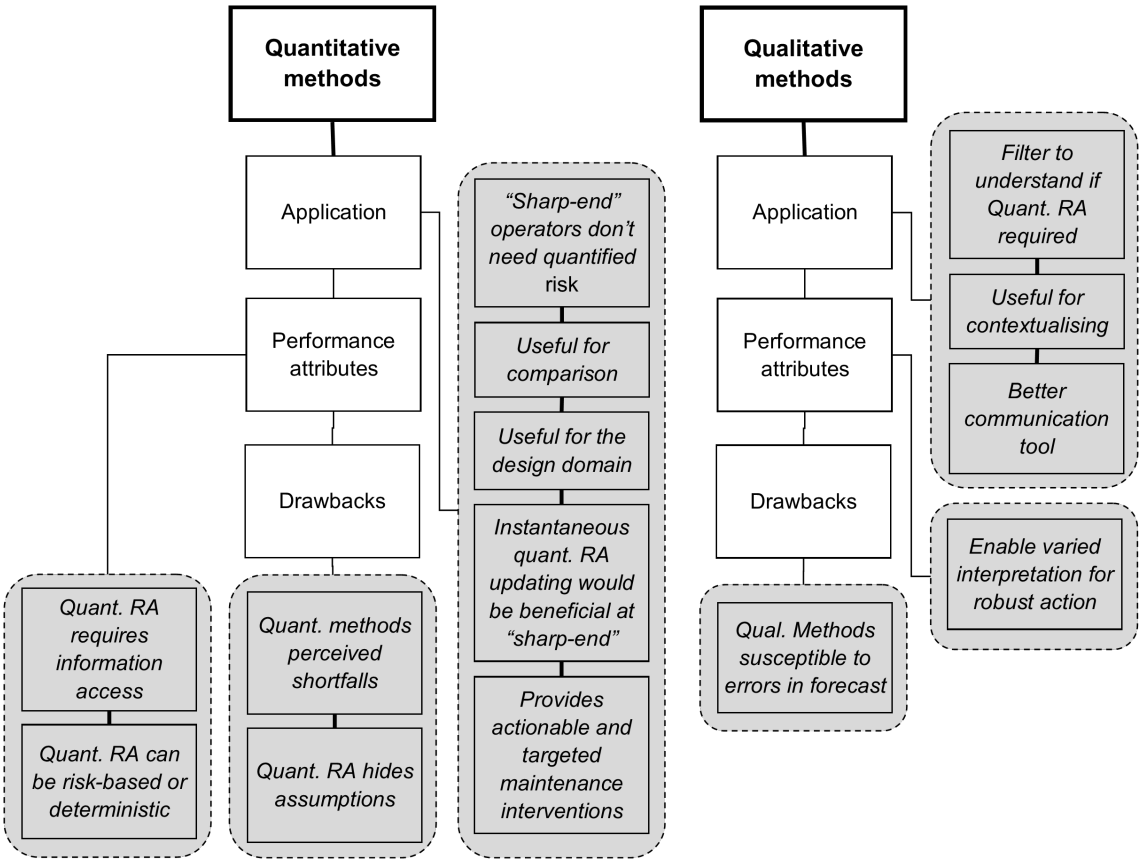
> *"If the practitioner is forced to assign a 'best guess' quantitative assessment for a risk which is not possible to accurately model and predict then this can be a very misleading representation to present to the risk owner."*

The availability of information was the most common factor conditional to the selection of either a quantitative or qualitative approach, although there was no common selection criterion identifiable. A consistent response was that both methods were mutually supportive, for example Participant 50 (aviation, designer/equipment manufacturer, risk analyst) stated:

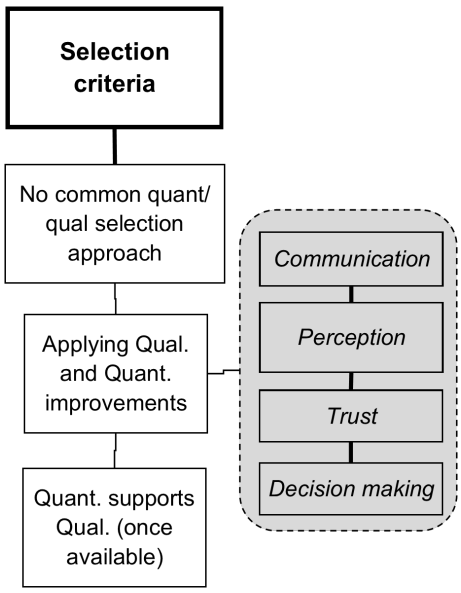
> *"We follow the qualitative approach invariably, because we do not have the scientific evidence to approach the probabilistic approach."*

Question 9 - "When handling in-service structural integrity risks, are you required to assess structural integrity risk relative to a threshold level?" There is a split of opinion with no statistically significant differences across predictor variables. The modal response from this sample was for "yes" (54.1% (20/37) $\pm$ 16.1% (95% confidence interval)) with "no" receiving 45.9% (17/37) $\pm$ 16.1% (95% confidence interval) responses, there is no significant difference at the 5% level. A weak significant relationship at the 10% level was identified between Question 9 and Question 8[3] (two-tailed Mann-Whitney U Test, $W = 218.0, p = 0.076$) - "no"

[3]Question 8 - "Within your organisation, are structural integrity risks that arise in-service described numerically or non-numerically/verbally/descriptively?"



(a) Themes relating to quantitative and qualitative method properties.



(b) Themes relating to criteria for selecting either quantitative or qualitative approach.

Figure 4.6: Survey Question 8e thematic analysis.

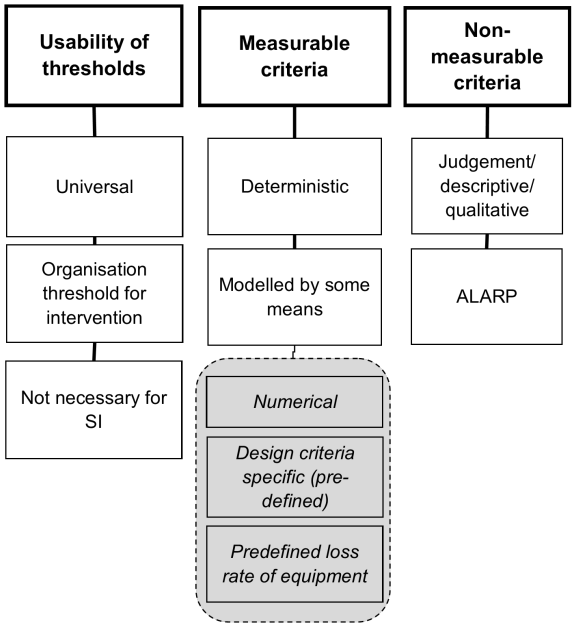


Figure 4.7: Survey Question 9a responses thematic map.

responses were more likely to answer "verbally" to Question 8. This may be a result of: the culture in the organisation being aligned to working with verbal representations of risk; of the individual being mistaken about the risk requirements within their organisation; or the organisation policy being incompatible with the risk representations seen by respondents.

Question 9a - "What is the nature of the threshold level? For example, your threshold level might be communicated as a loss rate per flying hour." *(Only "yes" responses from Question 9.)* Themes were identified regarding: the usability of thresholds; measurable criteria; non-measurable criteria; and answers that were not connected. Figure 4.7 shows the sub-themes, but notably, qualitative criteria were identified only by operator/end user and designer/manufacturer cohorts - regulator/policy maker participants made no reference to qualitative criteria.

Question 9b - "In operational situations which are information poor, time critical, or resource constrained, I believe a threshold level is appropriate." *(Only "yes" responses from Question 9.)* Organisation requirements were

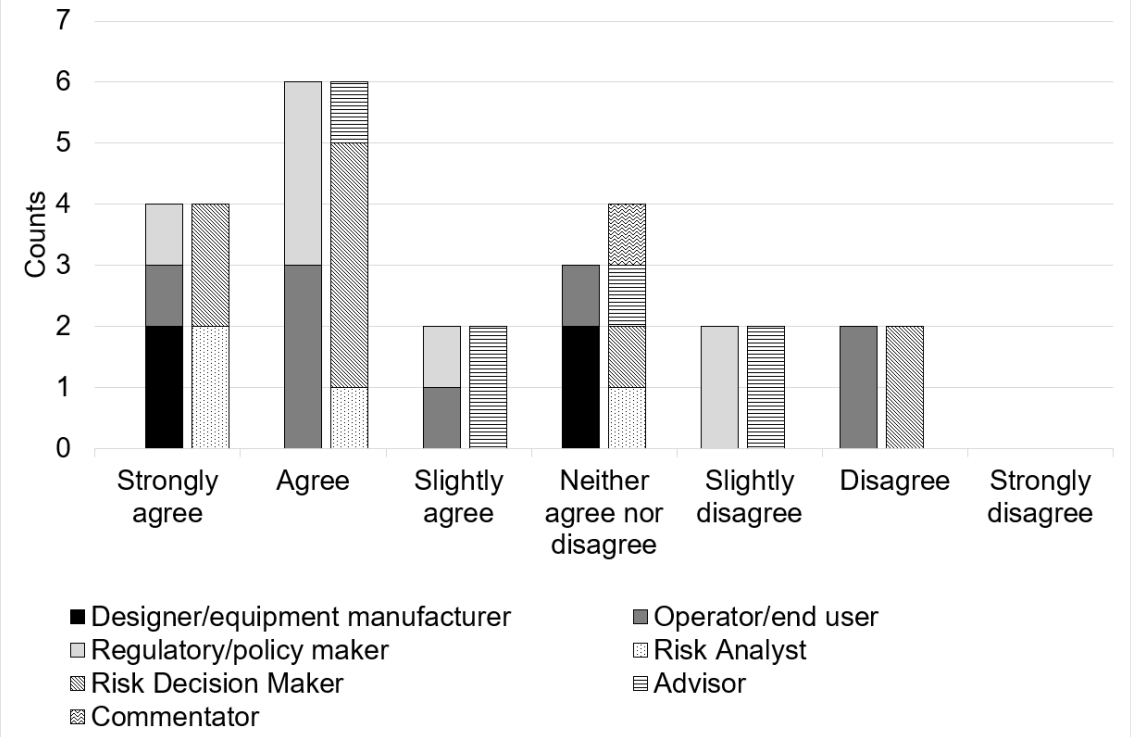


Figure 4.8: Survey Question 9b responses plot, grouped by respondent categories and roles.

expected to create a difference in response, however Figure 4.8 highlights variation in opinion across respondent roles and categories[4]. Notably, the designer/equipment manufacturer category all answered with no disagreement, while risk decision-maker roles showed the largest range, between "strongly agree" and "disagree".

Question 9c - "How is this threshold level set?" *(Only "yes" responses from Question 9.)* Across all respondents the modal response was "regulator" followed by "designer/manufacturer limits" (Figure 4.9). This variation might be expected across domains and roles, particularly given the differences in opinion regarding structural integrity risk already seen. However, separating results across roles provided a richer perspective, indicating that even within particular cohorts there are inconsistencies. For instance, the "regulator/policy maker" group indicated that threshold levels originate from either the regulator or

[4] "Neither agree nor disagree" indicates a difference in counts, because Participant 43 did not provide a category in response to Question 2.

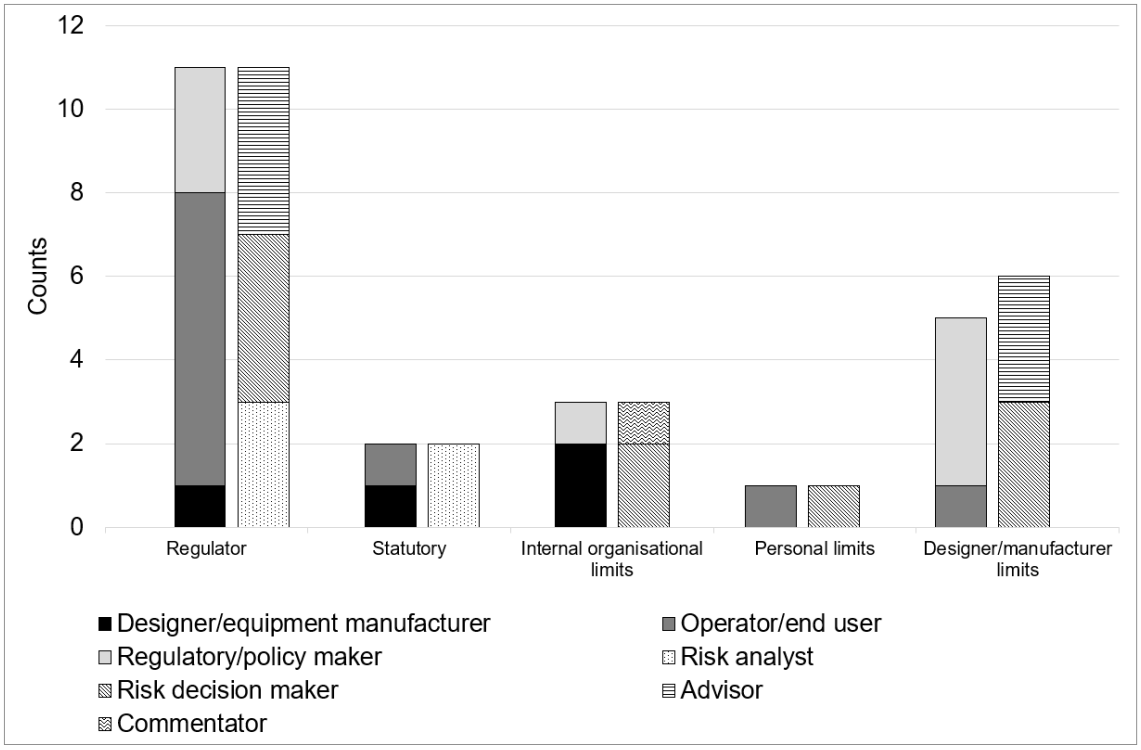


Figure 4.9: Survey Question 9c responses plot.

the designer/manufacturer.

Question 9d - "What processes do you use in support of structural integrity risk analysis or decision-making (tick all that apply)?" *(Only "yes" responses from Question 9.)* All methods offered to respondents received hits, but the most popular provide perspective of those methods that pervade most widely within the population. Specific areas may have dedicated methods, however the common methods are of interest in understanding practices and behaviours that are used to address structural integrity risk. The top 3 methods (including ties) for each predictor variable are shown in Table 4.1.

Question 9d1 - "Do you believe these processes are appropriate for structural integrity risk analysis in your organisation?" Participants were overwhelmingly in favour of the methods that they stated to apply: "yes" received 90% (18/20) of the responses.

		Des/eqpmt man.		Op/end user		Reg/pol maker		Risk analyst		Risk d-m		Commentator		Advisor	
	Remarks	n = 4		n = 8		n = 7		n = 4		n = 9		n = 1		n = 6	
	Other (specify)														
Stochastic processes	Neural Networks											1			
	Bespoke numerical models	3						2				1			
	PDfkV.	3						2				1			
	Bayesian analysis	3						3				1			
	MCS	2						1				1			
Hybrid	Bayesian Networks			3								1			
	Epistemic models											1			
	Risk Based Maintenance														
	Event tree analysis	3				3		2				1			
	Fault tree analysis	2		3		3		2		2		1		3	
	HEA technique							3							
Classical statistical approach	Event data models	1				3		1		3		1		3	
	Probability distributions	2						1				1		3	
Quantitative	QRA	1		2		1		1		1		1		1	
	Risk Matrix	2		3		1		2		2		1		1	
	PRAT							3							
	Bow Tie	3		2		3		2		2		1		3	
Qualitative	Task Audits	3						2				1		3	
	Safety Audits	2		1		2		1		1		1		2	
	Check lists			3		3		2						2	
	What If analysis	3		1				2		2				2	

Table 4.1: Top 3 risk analysis processes used to support structural integrity risk analysis and decision-making for respondent roles and categories. 1, 2 and 3 indicate the top, second and third most common answers given by that marginal group. PRAT: Proportional Risk Assessment Technique. HEA: Human Error Analysis. QRA: Quantitative Risk Analysis. MCS: Monte Carlo Simulation. PDfkV: Probabilistic distributions for key variables.

Question 9d1a - "Why do you believe this to be the case?" The themes emerging identify opinions, behaviours and functional aspects of risk analysis methods. Although overwhelmingly there was support, the themes undermine the credibility of how the approaches used are being applied.

- *Belief in RA Methods.* A common response was that the absence of historical structural integrity problems provides support for existing methods. This shows bias, perhaps in the form of a gamblers fallacy (Montibeller and von Winterfeldt, 2015, p1236), or a hasty generalization (Fulkerson, 1996, p99).
- *Information usage.* Judgement is used in the application of available information to risk methods, but while designers/equipment manufacturers desired as much information as necessary to conduct quantitative assessment, restricted access can limit the methods available to operator/end users. Participant 45 (aviation, regulatory/policy maker, advisor), who answered "no" to Question 9d1, commented specifically on contractual limitations restricting the completeness of the information set, with specific regard to non-destructive testing (NDT) records:

  > *"NDT database only contains data on NDT carried out by MOD not contractors. It is a useful but potentially incomplete picture. There is also no clear guidance for Structural EAs on how the data can be used to support SI management. NDT schedule reviews across various platforms are inconsistent, frequently cursory or sometimes non-existent. "*

  There was a spread of opinion about the use of historical evidence, conflicting between whether frequentist fault occurrence data is applicable for quantitative assessment or irrelevant. Participant 4 (aviation, operator/end user, risk decision-maker) answered "no", and commented on the methods applied:

  > *"They are suitable for identifying failure sequences but the techniques used to develop quantitative assessments are based*

> *on historical failure rates and do not assist with predicting the development of new unexpected in-service structural failures which were not expected within the design life of the platform."*

– *Functional factors.* Respondents found methods were functionally useful if they provided a clear structure to follow, enabled a clear audit trail for open challenge to analyses, and a consistency in approach. Operators and regulators preferred qualitative methods for failure sequence identification and to manage deviations from planned risk levels. Some respondents took the opportunity to enforce their belief that structural integrity is deterministic, and risk is created by incorrect design assumptions, or wrong consequence assessments. One participant's (operator/end user) belief that probabilistic accuracy required improvement, was complemented by a regulator/policy maker's belief that probabilities are only appropriate for demonstrating design standard compliance.

– *Situational factors.* Intensive methods requiring computation and simulation, more often in the designer/equipment manufacturer domain, were stated not to be usable in operating environments, because they were not "reasonably practicable". Operator/end users tended toward diagrammatic methods (bow-tie, event tree) that are less intensive.

Question 9e - "What measure/test is used to demonstrate a safe operating level in your organisation?" *(Only "no" responses from Question 9.)* 23.5% (4/17) participants in this sample stated they use no measure or test to demonstrate a safe operating level. For the remaining 76% (13/17), Figure 4.10 indicates the linking themes. Notably, the responses indicate reliance upon either pre-established measures (design specified or procedural), or they were adaptive to the situation and individual.

Question 10 - "Is your organisation's risk analysis methodology proactive, enabling anticipation of risk from in-use changes or modifications to

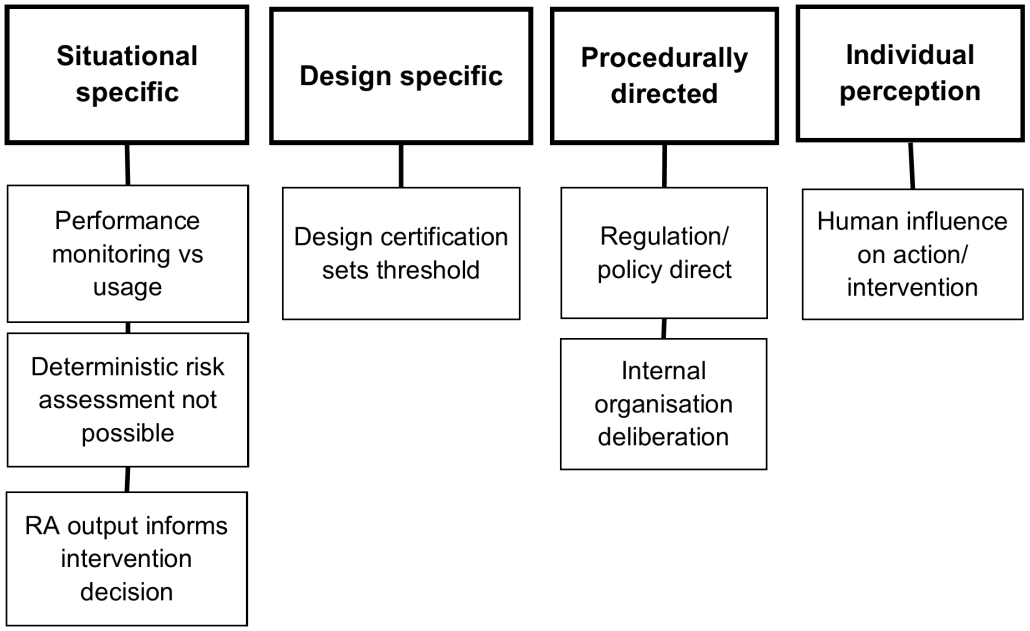


Figure 4.10: Survey Question 9e thematic analysis results.

usage/configuration, or is it reactive, simply responding to issues that occur?" Both "reactive" and "proactive" received 50% (19/38) $\pm$ 15.9% (95% confidence interval). There were no significant differences within the confounding factors (Mann-whitney non-parametric test). If differences had existed, this would have been unsurprising, with an example expectation being that design/manufacturers would act more proactively. However, the lack of difference across participants suggests that these answers reflect the experiences of the respondents in how they have seen risk approaches applied to structural risks.

Question 10a - "What methods do you use to identify risks?" Reactive methods tended to identify risks from human reporting or fault data. Proactive methods of risk identification were either prompted by procedural initiation, initiated from design approaches that provide practitioners the means to anticipate failures in the active flying fleet (for aviation respondents), or human generated.

Question 11 - "The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'likelihood' of occurrence." The modal response was joint "slightly agree" and "neither agree nor disagree" (22.2

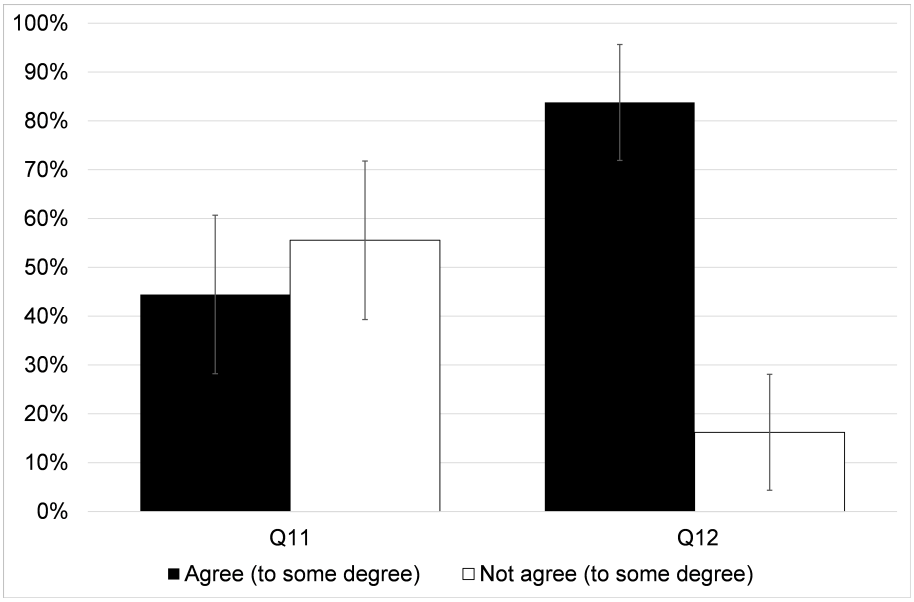


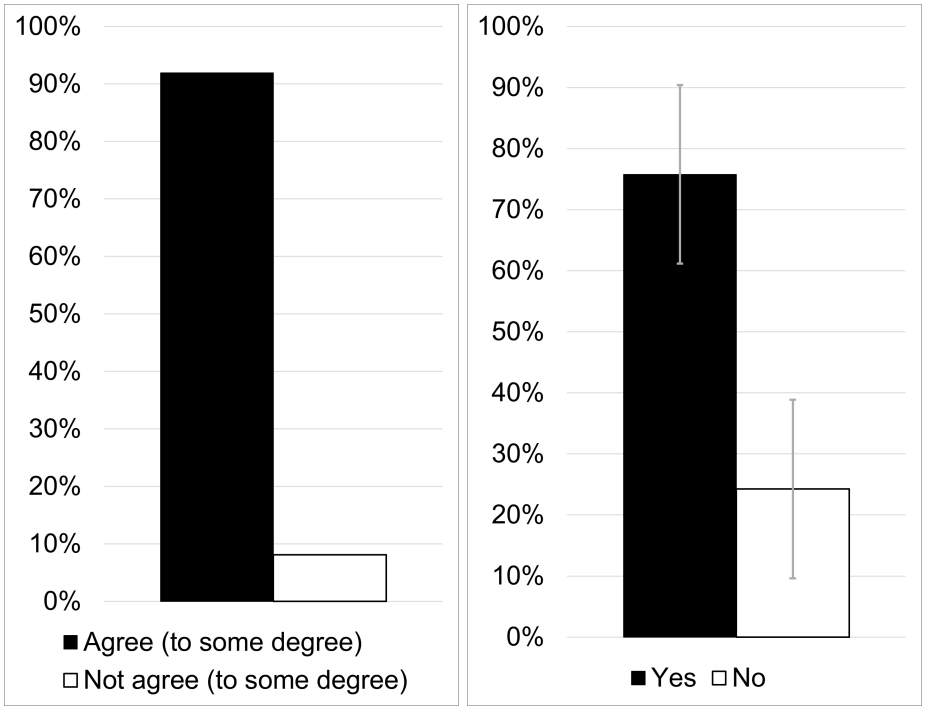
Figure 4.11: Questions 11 and 12 aggregated responses: The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'likelihood'(Question 11) / 'consequence' (Question 12) of occurrence.

(8/36) $\pm$ 13.6% (95% confidence interval)). Aggregating and comparing answers that agreed (to some degree) with answers that did not agree (to some degree), the latter of which included the central responses of "neither agree nor disagree", showed that there was a stronger preference within the sample towards not agreeing with the statement - 55.5% (20/36) not agreeing (to some degree) $\pm$ 16.2% (95% confidence interval), shown in Figure 4.11. No statistically significant differences were identified between category or role participant factors.

Of particular note is the finding that 12 respondents who answered "yes" to Question 9[5] did not agree (to some extent) that a clear likelihood is generated.

The inference from this data is that a large proportion of the population of structural integrity risk practitioners do not necessarily agree that a likelihood is able to be generated. This opinion may reflect upon the choice of approach taken when handling structural integrity risks, conformance with procedural requirements, individual attitudes toward risk management, behaviours when communicating risks, and consequences for those acting upon risk assessments for which likelihoods could not be meaningfully attributed.

[5]Question 9 - "When handling in-service structural integrity risks, are you required to assess structural integrity risk relative to a threshold level?"



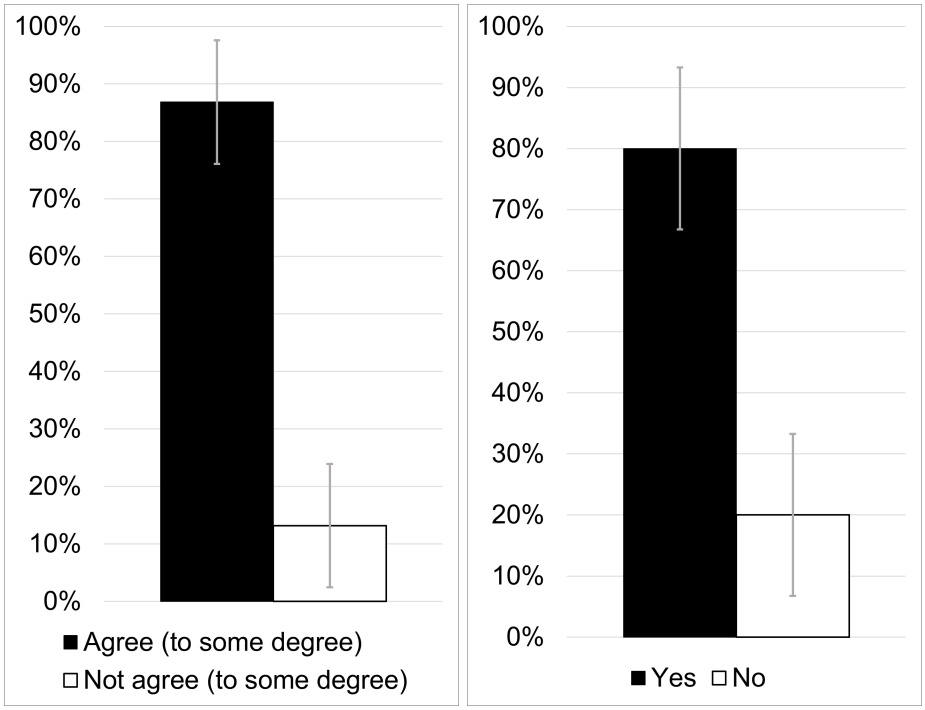
(a) Question 13 - "More often than not, there are uncertainties within my analyses of structural integrity risks."

(b) Question 13a - "Are uncertainties communicated in risk analyses regarding operational structural integrity?"

Figure 4.12: Survey Question 13 and 13a results analysis.

Question 12 - "The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'consequence' of occurrence." The modal response was "agree" with 46.0% (17/37) $\pm$ 16.1% (95% confidence interval). In the aggregated results, a clear majority within the sample was apparent (Figure 4.11) - 83.8% (31/37) agree (to some extent) $\pm$ 11.9% (95% confidence interval). The pattern of "not agree (to some degree)" was consistent across respondent roles and categories.

Question 13 - "More often than not, there are uncertainties within my analyses of structural integrity risks." The modal response was "agree" with 48.7% (18/37) $\pm$ 16.1% (95% confidence interval), with aggregated results indicating 91.9% (34/37) agreeing (to some extent) (Figure 4.12a). This strong agreement from the sample was consistent across respondent role and categories, and indicates that practitioners believe that structural integrity risk



(a) Question 14 - "More often than not, there are assumptions within my analyses of structural integrity risks."

(b) Question 14a - "Are assumptions communicated in risk analyses regarding operational structural integrity?"

Figure 4.13: Survey Question 14 and 14a results analysis.

analyses contain uncertainties. Notably, 70% of "yes" respondents to Question 9 (threshold assessment) agreed (to some extent) with Question 13, which indicates that uncertainties are included within comparisons of risk acceptability against threshold values.

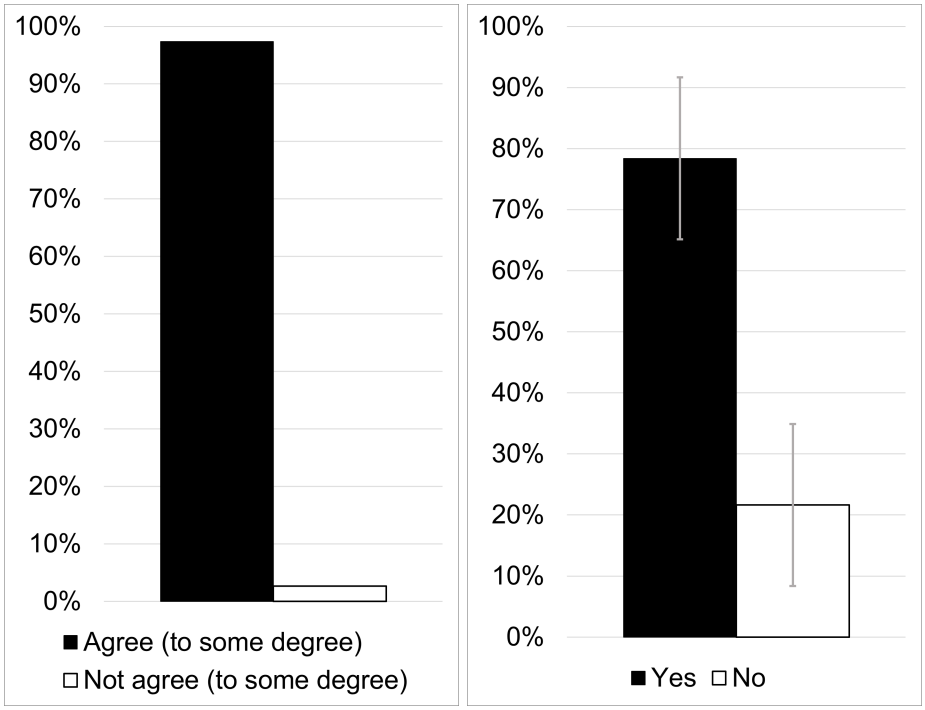
Question 13a - "Are uncertainties communicated in risk analyses regarding operational structural integrity?" *(Only "Neither" to "Strongly Agree" (inclusive) responses from Question 13.)* Although there are relatively wide confidence intervals, 24.2% (8/33) $\pm$ 14.6% (95% confidence interval) of respondents do not communicate the uncertainties contained in risk analyses (Figure 4.12b). This indicates that a sizeable proportion of the population do not communicate where there are uncertainties contained within risk analyses. All categories and roles (excluding commentator ($n = 1$)) were represented in the "no" response and no significant differences were identified.

Question 14 - "More often than not, there are assumptions within my analyses of structural integrity risks." The modal response was "agree" with 57.9% (22/38) $\pm$ 15.7% (95% confidence interval). Similar to Question 13, the aggregated results were heavily skewed to "agree (to some degree)" at 86.8% (33/38) $\pm$ 10.7% (95% confidence interval) (Figure 4.13a). Amongst role groups, a statistically significant difference was found when comparing risk decision-makers with non-risk decision-makers (two-tailed Mann-Whitney non-parametric test, $W = 242.5, p = 0.042$), with all "disagree" responses originating from risk decision-makers.

Responses conditional on Question 9 identified a similar outcome as for Question 13. In this case, 85% of "yes" respondents to Question 9 (threshold assessment) agreed that structural integrity risk analyses contain assumptions. This indicates that assumptions are included in assessments of risk acceptability against threshold values.

Question 14a - "Are assumptions communicated in risk analyses regarding operational structural integrity?" *(Only "Neither" to "Strongly Agree" (inclusive) responses from Question 14.)* A similar proportion of response to Question 13a was apparent, with 20% (7/35) $\pm$ 13.3% (95% confidence interval) of respondents answering "no" (Figure 4.13b). No significant differences were found when comparing respondent categories or roles.Of the 7 respondents who answered "no" to Question 14a, 3 had also answered "no" to Question 13a. This provides confidence that respondents were forming considered responses and not confusing assumptions and uncertainties.

Question 15 - "More often than not, I handle structural integrity risks conservatively." The modal response was "agree" (65.8% $\pm$ 15.1% (95% confidence interval)). An overwhelming majority (97.4%, 37/38) agree (to some extent) that conservatism is applied to structural integrity risk analyses (Figure 4.14a). This answer has ramifications when considering risk analyses against thresholds, or indeed measures of acceptability.



(a) Question 15 - "More often than not, I handle structural integrity risks conservatively."

(b) Question 15a - "When conservatism is applied, is it communicated in risk analyses regarding operational structural integrity?"

Figure 4.14: Survey Question 15 and 15a results analysis.

Question 15a - "When conservatism is applied, is it communicated in risk analyses regarding operational structural integrity?" *(Only "Neither" to "Strongly Agree" (inclusive) responses from Question 15.)* Compounding the use of conservatism, is the result that 21.6% (8/37) $\pm$ 13.3% (95% confidence interval) do not believe it is communicated when it has been applied to an assessment (Figure 4.14b). 40% (4/10) of regulator/policy makers believed this to be the case, while the remaining categories were more polarised. Within the role groups, all risk analysts answered "yes" ($n = 9$), advisors had 50% (5/10) "yes", and risk decision-makers had 82% (14/17) "yes". Irrespective, the sample size was insufficient for these results to be significant at the 5% level, and therefore representative of the population.

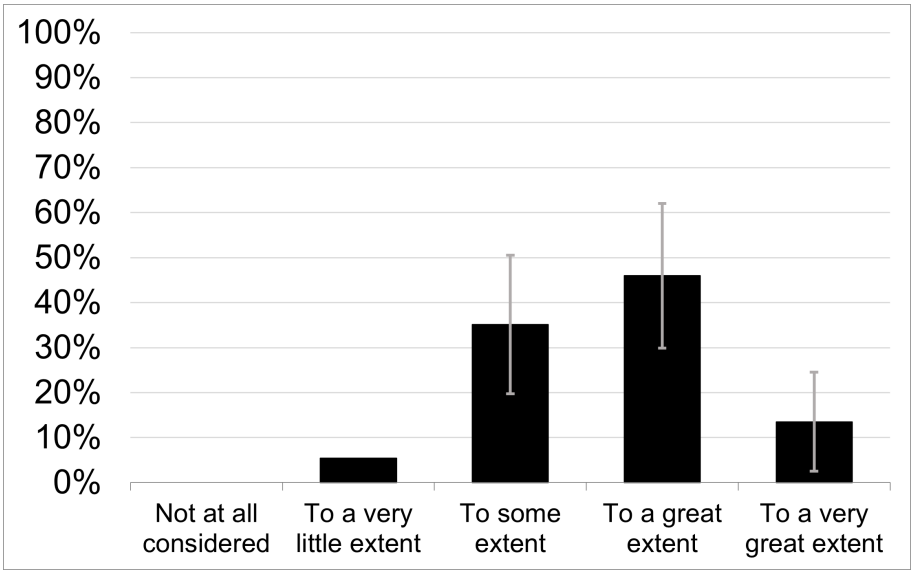
Question 16 - "To what extent do you believe human subjective judgement (from expert opinion or belief) influences the analysis of in-service structural integrity risks?" All respondents ($n = 37$) believed judgement had some influence upon the analysis of structural integrity risk (Figure 4.15a). The mode and median responses coincided around the likert category "to a great extent", with 59% (22/37) of all respondents believing that subjective judgement influences structural integrity risk analyses to either a great, or a very great extent. There were no statistically significant differences identified between respondent categories or roles.

Comparison with Question 9 indicates that 95% of respondents that answered "yes" to risk assessing relative to a threshold, answered that judgement influences structural integrity risk analyses "to some extent" or higher. Human subjective judgement therefore influences assessments made against prescribed thresholds.

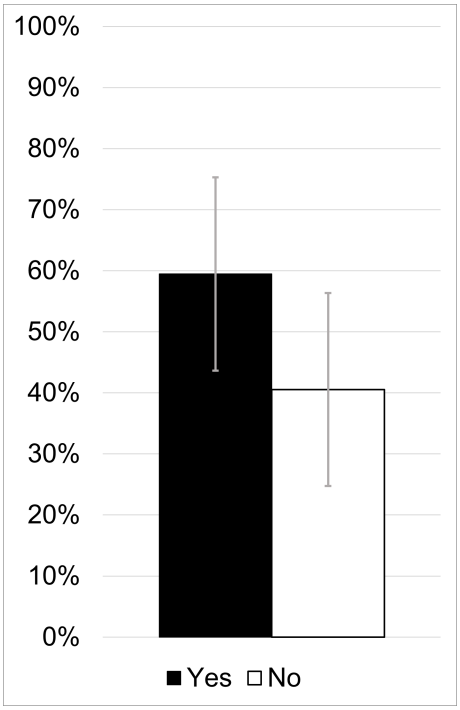
Question 16a - "Are human subjective judgements communicated in risk analyses regarding operational structural integrity?" At almost double the proportions given for Questions 13a, 14a and 15a, 40.5% (15/37) $\pm$ 15.8% (95% confidence interval) of respondents believed that human subjective judgements are not communicated in risk analyses (Figure 4.15b).

What might be perceived as differences in the sample between designer/equipment manufacturer respondents (87.5% "yes", 7/8) and others (48.1% "yes", 14/27) is not statistically significant at the 5% level (Fisher's exact method, $Z = 2.60, p = 0.101$). Operator/end users responses were skewed toward "no" (55.6%, 10/18), although this difference was also not significant at the 5% level compared with non-operator/end users (Fisher's exact method, $Z = 1.80, p = 0.106$). The high proportions of "no" responses infers that judgement-use is influential in structural integrity risk analyses, but stakeholders may be unaware of this influence and may be mislead by interpretive differences.

Question 17 - "I am confident in the structural integrity risk analysis method applied for handling operational risks." The modal response was "agree" with



(a) Question 16 - “To what extent do you believe human subjective judgement (from expert opinion or belief) influences the analysis of in-service structural integrity risks?”



(b) Question 16a - “Are human subjective judgements communicated in risk analyses regarding operational structural integrity?”

Figure 4.15: Survey Question 16 and 16a results analysis.

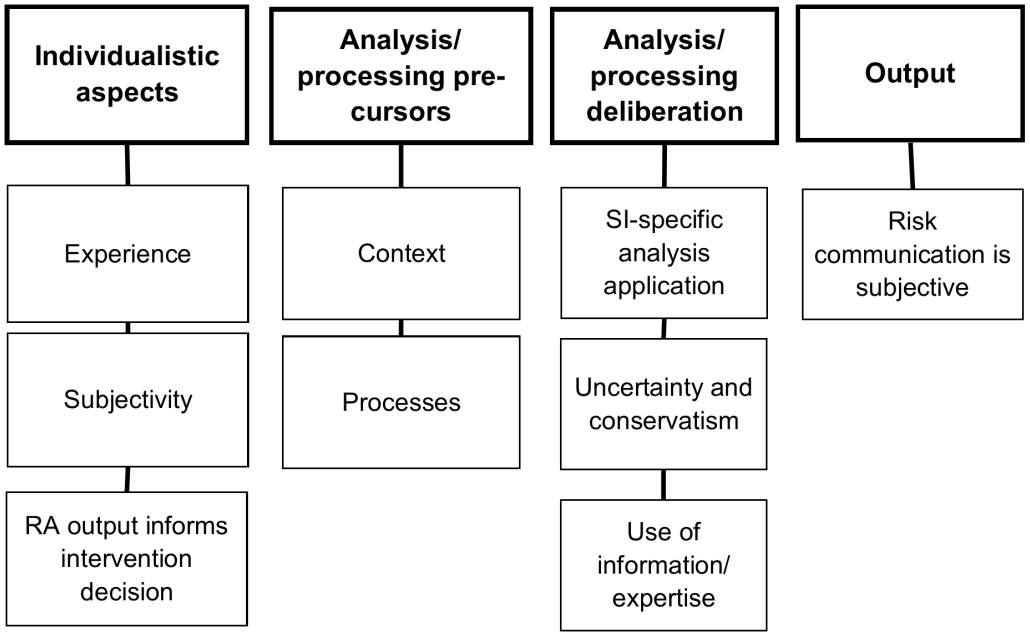


Figure 4.16: Survey Question 18 thematic analysis results.

50% (19/38) $\pm$ 15.9% (95% confidence interval). Respondents answered with 86.8% (33/38) $\pm$ 10.8% (95% confidence interval) agree (to some degree). There were no significant trends identifiable with respect to respondent category or role.

Question 18 - "Do you have any comments on the answers provided in this section?" Thematic analyses of responses highlighted reasserted individualistic positions on aspects of structural integrity risk analyses being dependent upon experience and subjectivity. The situational context frames the risk analysis, and there are pre-established perceptions about the subjective nature of risk when communicating analysis outputs. Figure 4.16 shows the thematic map.

Survey section 2 summary Section 2 provided insight into participant analyses of structural integrity risk, the process elements of which have been thematically grouped (Question 6). Structural integrity risk analysis is generally a valued activity (Question 7), however opinion varies regarding risk being presented qualitatively or quantitatively (Question 8). Questions 9 to 16 offered some insight into these perspectives, since organisations and individuals differed in opinion on whether they were required to apply threshold criteria for risk acceptability (Question 9), on

the processes they use in measuring structural integrity risk (Question 9d and 9e), and on the generation of risk-related parameters for analysing in-service structural integrity risk (Questions 11 to 16). For example, more respondents felt that a clear likelihood was not able to be generated, and all participants felt human subjectivity had influence on in-service structural integrity risk analysis.

4.1.3 Section 3 - Decision-making concerning structural integrity risk.

Question 19 - "What information is essential in order to make a decision upon an in-service structural integrity risk?" Thematic analysis identified the major themes as: environmental factors; practical solution constraints; measurable information; individualistic factors; and risk aspects (Figure 4.17). The sub-themes relate not only to extrinsic factors like fault data or mitigation options available, but also to intrinsic factors such as perceptions and metacognitive signals for example the "need for advice" and the "awareness of information state" (Cohen et al., 1996). The latter themes illustrate that internal processing of the situation is influential in the treatment of risk analyses.

Question 20 - "Describe the decision-making process regarding in-service structural integrity risks." Thematic analysis identified two clear sets of themes present in the data relevant to the process used by practitioners during structural integrity risk decision-making. One set grouped together descriptions of the type of process followed by participants, the other the discrete elements considered within the decision-making process.

Decision-making process themes Responses appealed to particular risk decision-making paradigms, which were grouped as: risk formulation focussed (seeking the elements of risk analysis); risk formulation but operationally constrained; operationally focussed but risk informed (considerate of risk during decision-making); and operationally focussed but deterministically informed (no

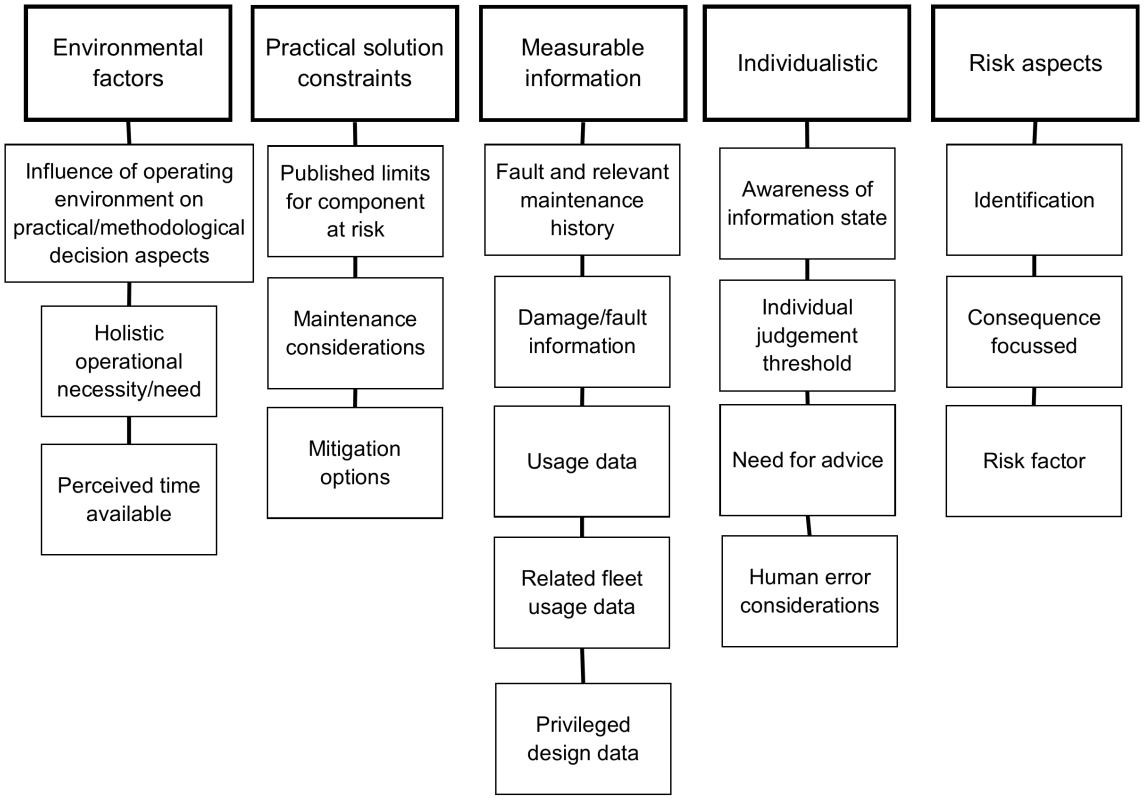
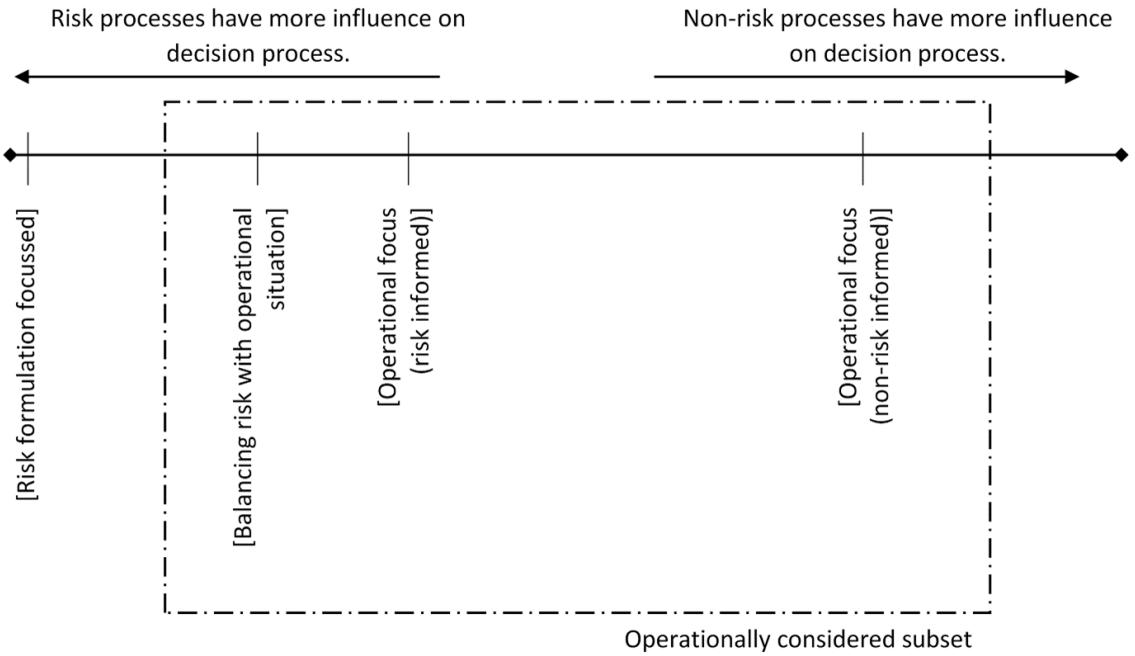


Figure 4.17: Survey Question 19 thematic analysis results.

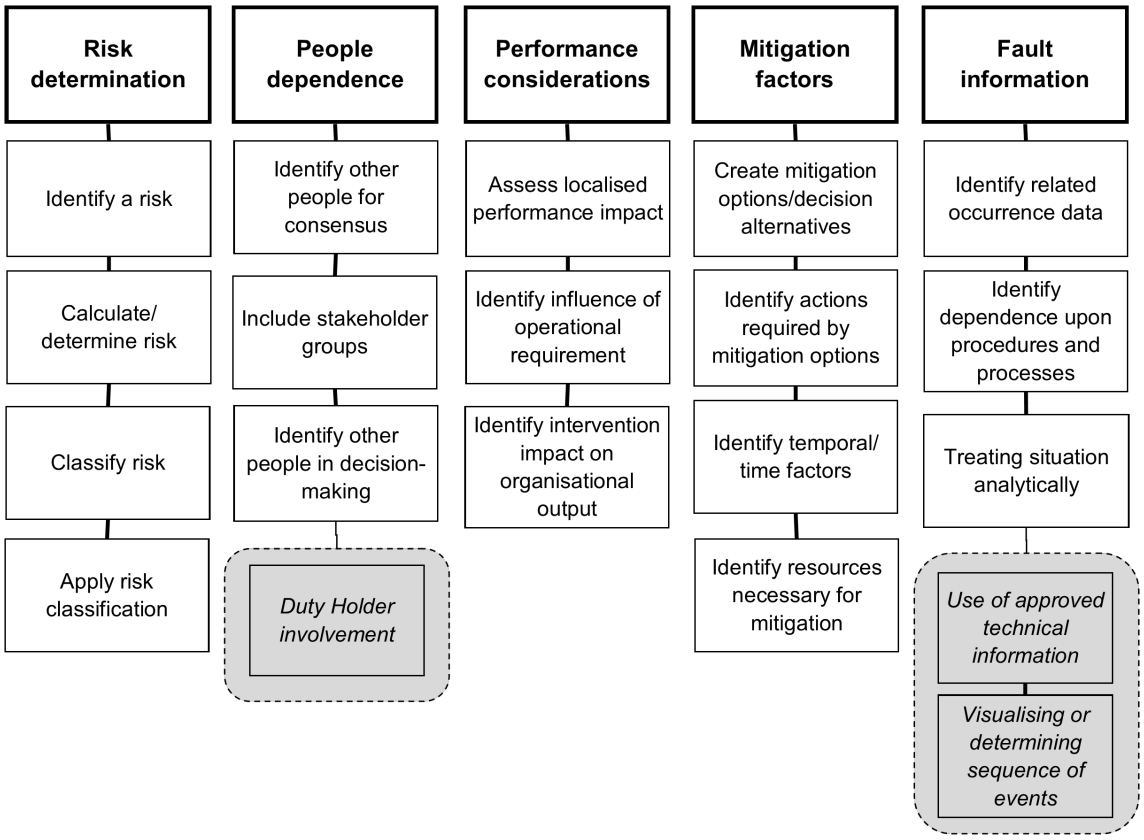
risk perspectives considered). Figure 4.18a describes the continuum between being risk-focussed and non-risk-focussed decision process themes, with a subset that identifies that includes operational factors.

Discrete decision-making themes Individual elements were grouped into activities (e.g. create mitigation options), behaviours (e.g. treating situation analytically) and considerations (e.g. performance considerations) made by respondents in forming their decision. These themes were grouped as reliant upon: determination of risk; having access to key people; considering performance implications of the decision problem; identifying the intervention and mitigation practicalities; and processing of fault information. Figure 4.18b includes the sub-themes.

Question 21 - Do you ever use judgement when making or supporting a decision regarding in-service structural integrity risks? Despite the



(a) Survey Question 20 - processes identified



(b) Survey Question 20 - decision elements identified

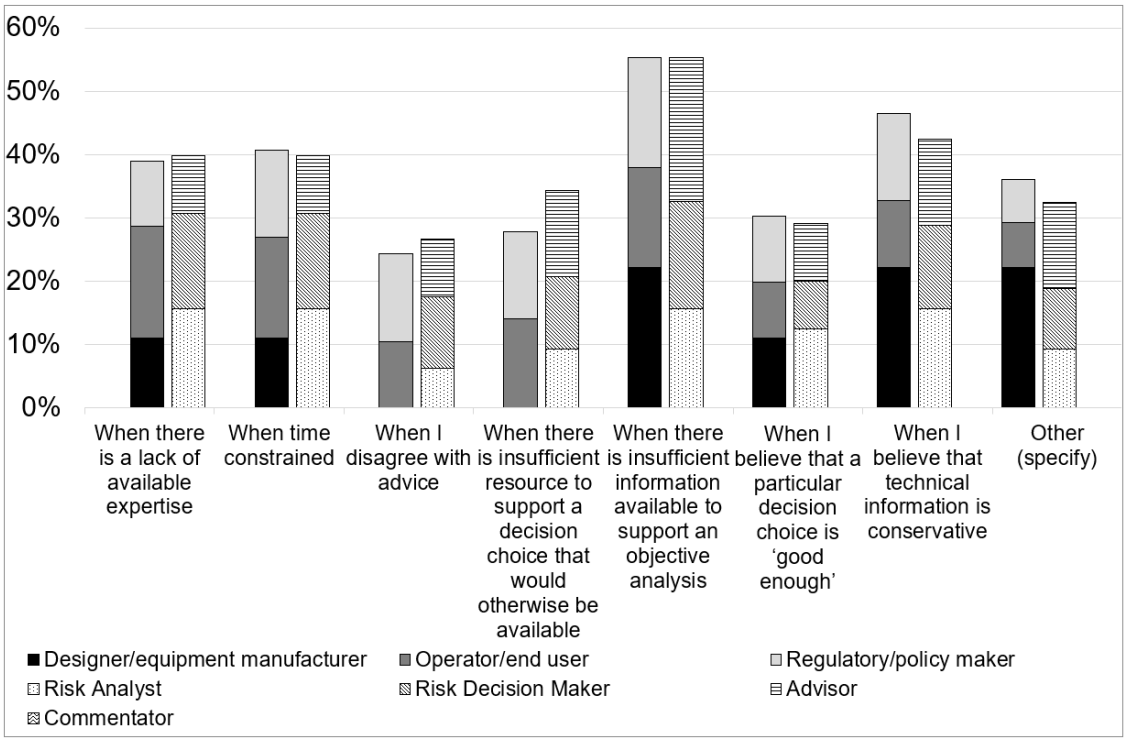
Figure 4.18: Survey Question 20 responses.

answers to Question 16 indicating the influence of judgement on structural integrity risk analyses, a minority of respondents answered that they do not use judgement: 13.9% (5/36) $\pm$ 11.3% (95% confidence interval). There were no significant differences identifiable between respondent role or category, although notably non-operator/end users were 4 times higher in responding "no" (not significant using a test of proportions (Fisher's exact method, $Z = -1.39, p = 0.340$)), and all risk analysts responded "yes".

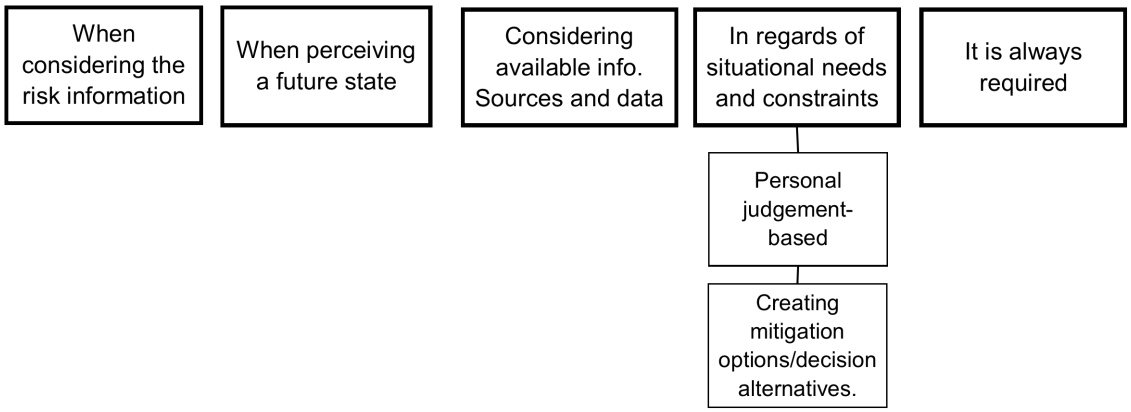
Question 21a - In which types of situation do you use judgement? *(Only "Yes" responses from Question 21.)* Multiple answers could be selected by participants, and all options presented to the participants received responses (Figure 4.19a). The modal response was "when there is insufficient information available to support an objective analysis" (65.5% (19/29) $\pm$ 17.3% (95% confidence interval)), followed by 3 options that all received 51.7% (15/29) $\pm$ 18.2% (95% confidence interval) of participant responses: "when there is a lack of available expertise", "when time constrained" and "when I believe that technical information is conservative".There were noticeable differences between the categories of respondents, particularly that no designer/equipment manufacturers responses were received for "when there is insufficient resource to support a decision choice that would otherwise be available" and "when I disagree with advice".

The 37.9% (11/29) $\pm$ 17.7% (95% confidence interval) of responses given as "other" were thematically analysed and the results are presented in Figure 4.19b. The responses reinforce the variety of states in which practitioners rely upon judgement.

Question 21b - "Is your judgement holistic (by gut, by intuition, or by instinct), or do you decompose your judgement upon the situation into constituent parts?" *(Only "Yes" responses from Question 21.)* 73.3% (22/30) $\pm$ 15.8% (95% confidence interval) of practitioners decompose their judgements, and although there were no significant differences across roles or categories,



(a) Survey Question 21a quantitative results.



(b) Survey Question 21a "other" responses thematic map.

Figure 4.19: Survey Question 21a responses.

non-operator/end users had a greater proportion of "holistic" responses.

Question 21b1 - "Are the judgements you make numerical or non-numerical / verbal?" The dominant response was "both" (66.7% (20/30) $\pm$ 16.9% (95% confidence interval)). Only risk analysts provided responses for "numerical" (6.7% (2/30)).

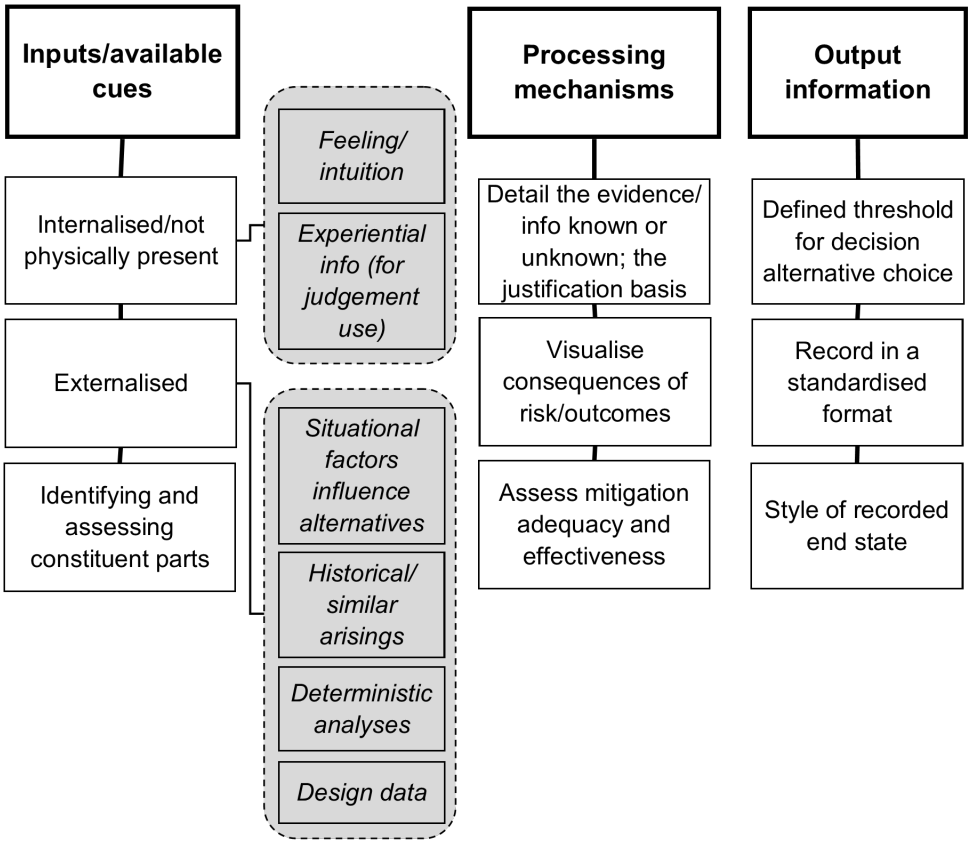


Figure 4.20: Survey Question 21b2 thematic analysis results.

Question 21b2 - "How do you represent the decomposed the situation?" *(Only "Decomposed" responses from Question 21b.)* For those respondents who stated they decompose their judgements, the themes identifiable were grouped into a structure of input, processing, and output. The input grouping in particular identified sub-themes that pooled elements particular to internalised value aspects, such as scenario urgency and experiences, and externalised aspects that are tangible such as analyses or the use of design data. Figure 4.20 shows the sub-themes.

Question 21c - "What do you find difficult when using judgement to support decisions regarding in-service structural integrity risk?" *(Only "Yes" responses from Question 21.)* The primary themes generated from Question 21b2 (input, processing, output) were used as a template to analyse the discursive responses. The individual elements demonstrated the difficulties experienced by individuals when applying their judgement, such as in novel situations, using the

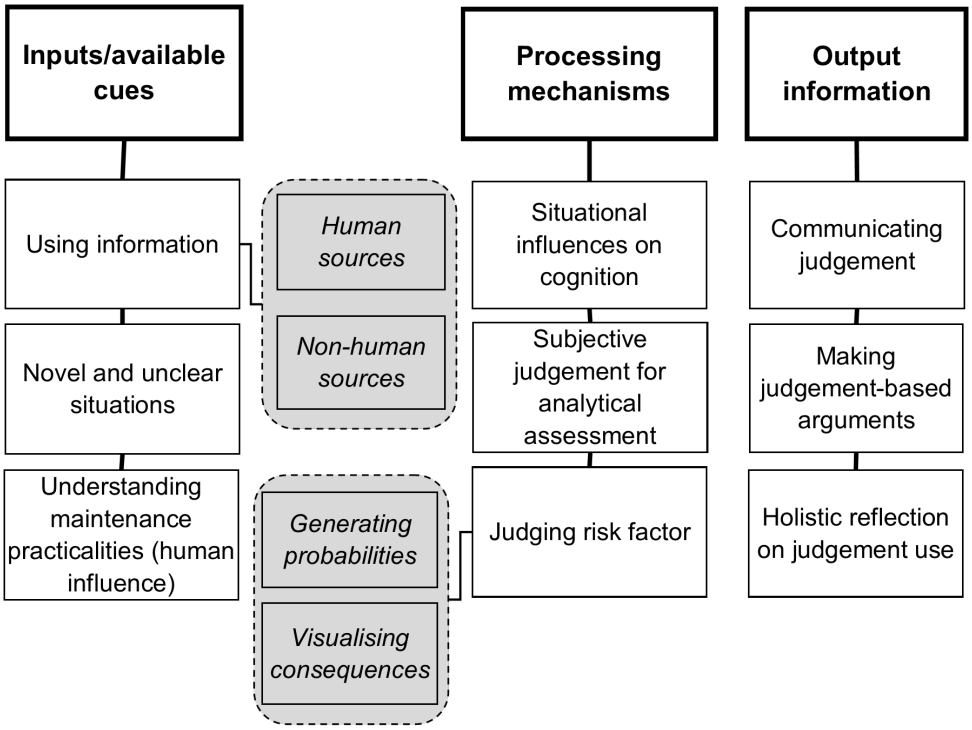


Figure 4.21: Survey Question 21c thematic analysis results.

information that is available (from human and non-human sources), in assessing situation-specific mitigations, and explaining the resulting judgement to others. These are consistent with the research project boundary conditions. Figure 4.21 shows all of the sub-themes identified.

Question 21d - "How could you be supported better when using judgement to support decisions regarding structural integrity risk?" *(Only "Yes" responses from Question 21.)* Although a self-reflective question, it was hoped this set of answers may provide direct steers to the most critical shortfalls in judgement-based decision-making regarding structural integrity risks. Three major themes connected the responses: knowledge, information and analysis. Although the responses appeared to be calling for methods to improve assessment efficacy through improved access to information, improved quantitative risk analysis generation, and access to specialist analysis, some answers called for more qualitative improvements. For instance, calling for greater value to be placed on corporate knowledge and peer review infers there is currently a lack of trust or appetite to use practitioner experiential knowledge. Similarly, some responses

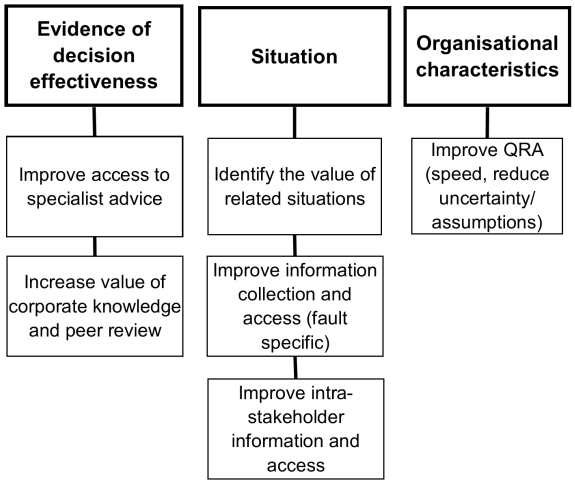


Figure 4.22: Survey Question 21d thematic analysis results.

stated that related situations were of importance to decision-making, which suggests that analogous knowledge is an accessible judgement-driven cue. Figure 4.22 shows these sub-themes.

Question 21e - "Have you received training in applying judgement to support decisions regarding structural integrity risk?" *(Only "Yes" responses from Question 21.)* A 50% (15/30) $\pm$ 17.9% (95% confidence interval) split masks a statistically significant difference between respondent groups. Operator/end users were significantly more likely to respond "yes" to having received training compared with non-operator/end users, shown by a test of difference between proportions that indicated a significant difference ($Z = -3.07, p = 0.002$). This result compares interestingly with Question 16, where non-operator/end users believed more strongly than operator/end users that risk analyses were influenced by human subjective judgements.

There were no significant differences between respondent roles, although within this sample, risk analysts responded with a larger proportion of "no" answers, and risk decision-makers had a larger proportion of "yes" responses.

Question 22 - "In operational situations, decisions regarding a structural integrity risk are clearly justified." There was a clear preference towards "agree (to any extent)", and no significant differences were identified between roles

or categories. The only "disagree" responses were received from 2 operator/end user risk decision-makers.

Question 22a - "Is there a mechanism for ensuring that structural integrity risk analyses are justified?" The modal response was "yes" (61.8 (21/34) % $\pm$ 16.3% (95% confidence interval)). There were no differences identified across respondent categories or role, with the split between "yes" and "no" being similar.

Question 22a1 - "What is the mechanism that you use for ensuring that structural integrity risk analyses are justified?" *(Only "Yes" responses from Question 22a.)* Thematic analysis identified the following themes, with no sub-themes: individual judgement; written record of assessment; policy requirement; peer review and internal scrutiny; and stakeholder working groups and meeting reviews. Again, the presence of individual judgement and a reliance upon other human deliberation through peer review or stakeholder meetings infers a dependence upon non-analytical means to handle structural integrity risk analyses.

Question 23 - "Are you more comfortable with numerical or non-numerical / verbal / descriptive representations of risky decisions regarding in-service structural integrity?" There was an even split in responses between "numerical" and "verbal", both proportions were 50% (18/36) $\pm$ 16.3% (95% confidence interval). Risk decision-makers showed a preference towards "verbal" representations of risky structural integrity decisions, although this difference with non-risk decision-makers was not significant at the 5% level (test of difference between two proportions, Fisher's exact method, $Z = 2.12, p = 0.094$).

Comment is necessary on the differences between answers to Question 23 and similar questions earlier in the survey. The "both" responses in Questions 8 and 21b1 were almost evenly distributed between numerical and verbal choices in Question 23. However, the comments from Question 18 suggest that this

is not necessarily as polarised as the results suggest, with qualitative and quantitative forms considered to be complementary and to provide a richer basis for decision-making. No "both" option was offered in Question 23, unlike in Questions 8 and 21b1. This was done to provide insight for future framework development to support decision-making about structural integrity risk.

Question 24 - "Within your organisation, are decisions upon structural integrity risk assessments "risk-informed" or "risk-based"?" The definition of risk-informed presented to the participants was: "*Risk assessment results support decision-making but are not the only consideration*". The definition of risk-based presented was: "*Risk assessment results form the sole basis for decision-making*". The modal response was "risk-informed" (79.4% (27/34) $\pm$ 13.6% (95% confidence interval)). There was a noticeable difference in proportions between risk decision-makers (94.1% (16/17) "risk-informed"), and non-risk decision-makers (64.7% (11/17) "risk-informed"), however the difference was not significant at the 5% level (test for two proportions, Fisher's exact method, $Z = -2.28, p = 0.085$). No differences were identified between respondent categories.

Question 25 - "In operational situations, have you received pressure or influencing force to make a decision on a structural integrity risk?" The modal response was "yes" (72.2% $\pm$ 14.6% (95% confidence interval)), and a noticeable but not significant difference (at the 5% level) was found between advisors (44.4% (4/9) "yes") and non-advisors (81.5% (22/27) "yes") - test of difference between two proportions, Fisher's exact method, $Z = -2.04, p = 0.079$. Although not significant, the result is consistent with expectations about how advisors operate in professional settings - independent of internal organisation pressures and politics. No other differences were identified, which suggests that pressures are inherent to all categories of operational engineering risk practice.

Question 26 - "In your opinion, in-service structural integrity risk analyses are used effectively by Decision-Makers" The grouped mode was "agree to any extent" (86.1% $\pm$ 11.3% (95% confidence interval)), "agree" being the modal response across all options. No significant differences were identified across respondent categories or roles.

Question 26a - "Why do you believe this to be the case?" The themes connecting the responses for why participants believe decision-makers use risk analyses effectively identify some interesting perceptions about decision-maker risk behaviour (Figure 4.23). Firstly, believing risk analyses as leading to the correct decision may be evidence of outcome bias or halo effects: that a good outcome *must* mean a good decision was made based upon the risk analyses. Secondly, the situation compels a subjective deliberation, but subjectivity is also recognised to influence decision-maker performance. Thirdly, the practices for handling risk assessments within an organisation control the level of risk decision-makers are exposed to, but individual decision-makers can still be misled by their understanding of analyses.

Question 27 - "Considering situations that you have seen of structural integrity risk management decision-making, do you believe Decision-Makers understand the assumptions and uncertainties presented to them from structural integrity risk analyses?" For both questions, the modal response was "most of the time" (assumptions - 60.6% (20/33) $\pm$ 16.7% (95% confidence interval); uncertainties - 55.9% (19/34) $\pm$ 16.7% (95% confidence interval)), and an overwhelming majority responded "atleast half of the time" for both (assumptions - 90.9% (30/33) ; uncertainties - 82.4% (28/34) $\pm$ 12.4% (95% confidence interval)).

Question 27a - "List up to 3 most common reasons that Decision-Makers do not understand assumptions with structural integrity risk analysis?" *(Only "Never" to "Most of the time" (inclusive) responses from Question 27.)* The

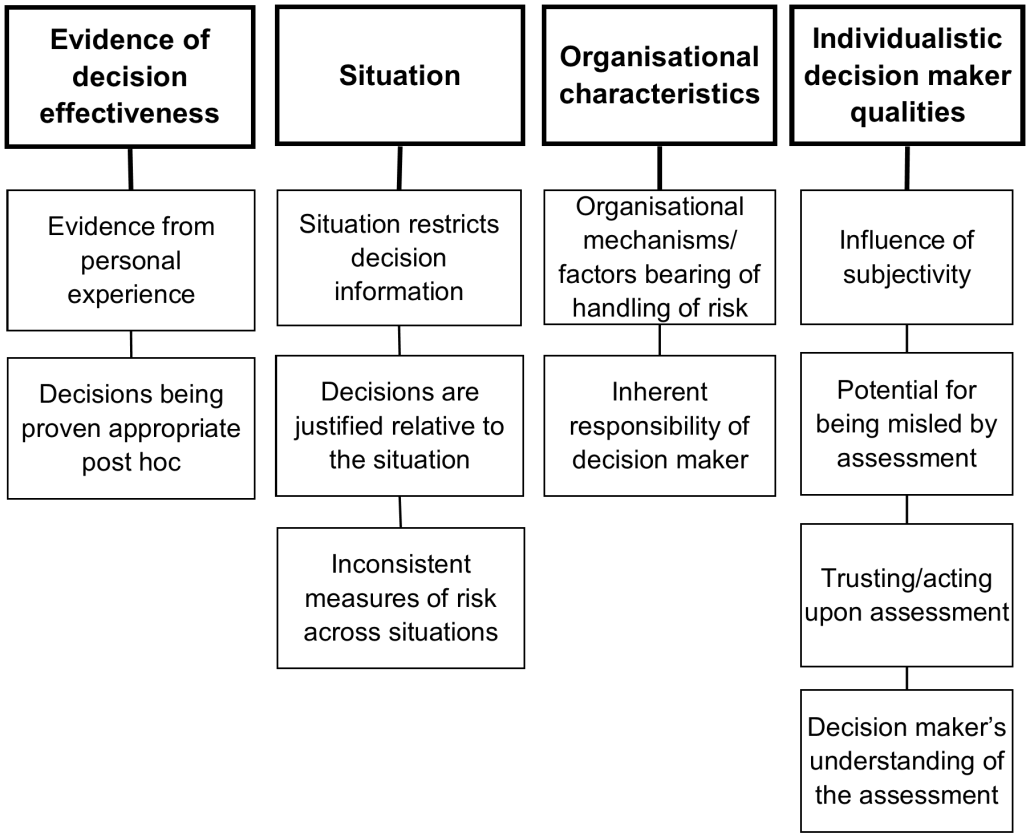


Figure 4.23: Survey Question 26a thematic analysis results.

main themes extracted related to decision-makers not understanding assumptions were:

- *Precursors.* The sub-themes related to this were: knowledge base; training deficiencies; knowledge of structural design; understanding the connection between operating and engineering risk; and structural complexities influencing engineering risk.
- *Forming the risk assessment.* The sub-themes related to this were: analytical and methodological restrictions on the analyst, decision-maker understanding of the risk assessment process; and the manner in which assumptions are described.
- *Communication and deliberation of risk assessment.* The sub-themes related to this were: situational and environmental pressures; and quality accuracy and adequacy of information.

– *Decision-maker rationale.* The subtheme related to this was the transparency of the rationale.

Question 27b - "List up to 3 most common reasons that Decision-Makers do not understand uncertainties within structural integrity risk analysis?" *(Only "Never" to "Most of the time" (inclusive) responses from Question 27.)* The main themes extracted related to decision-makers not understanding uncertainties were:

– *Situation and equipment-specific aspects.* The sub-themes related to this were: situational factors (urgency, or usage); and equipment usage.

– *Handling information.* The sub-themes related to this were: presenting uncertainties to decision-makers; an inability to work with uncertainty; and information features (quality, quantity, availability).

– *Decision-maker attributes.* The sub-themes related to this were: knowledge base; specific structural knowledge requirements; understanding risk assessment processes; and behavioural aspects.

Decision-maker cognition and behaviour was also common to understanding about uncertainties. The other themes considered that information presentation, uncertainty and completeness influenced decision-maker understanding, alongside situation-specific detail necessitating individualistic consideration about factors such as usage or urgency.

Survey section 3 summary The questions in section 3 elicited perspectives on decision-making about structural integrity risk. The information necessary (Question 19) and process used for decision-making (Question 20) highlighted the participant-perceived important elements and conditions necessary for decisions about structural integrity risk to be formed. Judgement and decomposition of the situation is applied by most respondents for decisions on risk (Question 21); notably, judgement and social deliberation also have bearing upon decision justification (Question 22). A number of situations necessitate the use of judgement

(Question 21a) and various difficulties surrounding the handling of structural integrity information for input, processing and output activities can create difficulties in applying judgement (Question 21c). These responses are compounded by most participants having received pressure to make a structural integrity risk decision (Question 25). Decision-maker cognition and behaviour were identified as common reasons for not understanding assumptions and uncertainties in structural integrity risk analyses, alongside risk communication, decision rationale and situational influences (Question 27).

4.1.4 Section 4 - Management of structural integrity risks arising in-service and handling of risk analyses conducted

Question 29 - "Were you previously, or have you currently been formally delegated the authority to own and decide upon structural integrity risks in your position?" Providing an additional perspective on the background of the participants, a majority of the sample participants have been delegated with structural integrity risk decision-making responsibilities (82.4% (28/34) $\pm$ 12.8% (95% confidence interval)). Regulator/policy makers participants showed a higher proportion of "no" responses (40.0% (6/10) $\pm$ 17.0% (95% confidence interval)) compared to non-regulator/policy makers (9.1%, 20/22), but the difference in proportions was not significant at the 5% level (Fisher's extact method, $Z = 1.86, p = 0.06$). A larger sample may confirm this interesting result, which infers that some practitioners in a position to influence ways of working may not be in a position to empathise with the decision-making context. It may also explain some of the different risk assessment methods used between regulator/policy makers and operator/end user groups (see Question 9d1). Unsurprisingly, risk decision-makers responded with the largest proportion of "yes" answers (94.1% (16/17)) although this was not a significant difference to non-risk decision-makers at the 5% level (test of difference between two proportions, Fisher's exact method, $Z = -1.89, p = 0.175$).

Question 29a - "Did/do you own this risk solely or in conjunction with others?" *(Only "Yes" responses from Question 29.)* The modal answer was "in conjunction with others" (60.7% (17/28) $\pm$ 18.1% (95% confidence interval)) and the proportions were relatively consistent across all respondent categories and roles except for advisors, who showed a preference toward "sole owner". This infers that decision-making responsibilities for risk are, largely, collaborative. Under these conditions, transparency of risk decision-making becomes a necessity in order to manage risk consistently. For example, an action to mitigate a risk in one environment may not transfer readily to a second environment, so the conditional requirements of the mitigation must be clear to the multiple responsible parties.

Question 30 - "Your organisation's risk management method enables you to maintain a transparent audit trail, which details the history of the treatment and ownership of in-service structural integrity risks." Overall, there was agreement (82.4% $\pm$ 12.8% (95% confidence interval)). Risk analysts responded with strong agreement and no "disagree" responses. The 5 "disagree (to any extent)" responses indicated were from operator/end user and regulator/policy maker categories, and the risk decision-maker and analyst categories.

Question 30a - "What measures/aspects of the risk management framework enable the process to be considered as transparent?" *(Only "Neither" to "Strongly Agree" (inclusive) responses from Question 30.)* Four key themes were identified by thematic analysis:

- *Procedures and Tools.* Pre-determined tools and procedures define the risk management process.
- *Stakeholder involvement.* Information sharing and discussions with stakeholders demonstrates transparency. Information is maintained in shared working environments to allow access to risk related material, although access is controlled for commercial or security reasons. Discretion is applied on who requires sight of risk information, which may be masked by structural

damage being within repair capability. Safety and risk-related meeting substance and outcomes are recorded.

– *Independent involvement.* External parties and assurance mechanisms provide advice and independent perspectives on risk management processes.

– *Recording risk related activity.* Technical queries, formal safety review activity, meetings, decisions made, and relevant evidence supporting risk management is recorded. Information is logged from operators, relating to risky occurrences, although issue reporting may be influenced by subjectivity or if any issue is repairable, which may skew trending data. Data recorded from meetings may miss the subtlety of judgements made.

Question 31 - "Who do you believe acts upon the risk analysis conducted for in-service structural integrity issues? (select all that apply)" The modal response was "operator/end user" (85.3% (29/34) $\pm$ 11.9% (95% confidence interval)) across all roles and categories, although manufacturer/equipment owner scored highly (70.6% (24/34) $\pm$ 15.3% (95% confidence interval)). Irrespective, all answer options received responses from all categories and roles, except for advisor (who recorded nil answers for "Another department internal to organisation") and commentator ($n = 1$) roles. This infers that practitioners believe multiple stakeholders act upon risk analyses conducted.

Question 32 - "To what extent does your structural integrity risk management strategy explicitly try to identify rare events, 'black swans' or 'unknown unknowns'?" The modal response was "to some extent" (45.5% (15/33) $\pm$ 17.0% (95% confidence interval)), with designer/manufacturers indicating a lower level (to a very little extent) compared to the other groups (to some extent). Respondent roles returned the same median (to some extent). The results infer that within the population of engineering risk practitioners, there is a minority proportion who do not at all consider structural integrity black swans

(18.8% (6/33) $\pm$ 13.2% (95% confidence interval)).

Question 32a - "Do you use a formalised technique to identify rare events?" *(Only "To a very little extent" to "To a very great extent" (inclusive) responses from Question 32.)* The modal answer was "no" (83.3% (20/24) $\pm$ 14.9% (95% confidence interval)), and for those that responded "yes", the answers given referred to organisation policy approaches (specifically the ESVRE framework mentioned in Chapter 2), or were vague in their justification:

> Participant 36: *"I don't know"*
> Participant 57: *"Risk tools"*

Question 33 - "Is Structural Health Monitoring planned, or currently in use on your equipment for use in Structural Integrity risk management?" The majority response was "yes" with a proportion of 78.8% (26/33) $\pm$ 14.0% (95% confidence interval) ("no" received 12.1% (4/33) $\pm$ 11.1% (95% confidence interval); "unsure" received 9.1% (3/33)). There was no significant pattern identifiable within the predictor variables, although 20.8% (5/24) $\pm$ 13.9% (95% confidence interval) respondents stated they believe the accuracy of the risk analyses is fair to very good (Question 8a), despite not knowing whether their equipment is using structural health monitoring. This infers that these respondents have strong belief in the methods applied for conducting risk analysis, or perhaps perceive risk accuracy more generally, than at the abstraction of individual issues.

Question 34 - "What methods are used to provide confidence in the risk analysis conducted (for example independent review, consensus of opinion, audit)?" The responses were grouped thematically as data-related methods, guiding principles (such as independence or assurance), and people-focussed aspects (such as by gaining confidence discussing with peers, experts or generalists). Figure 4.24 shows the sub-themes for each of these.

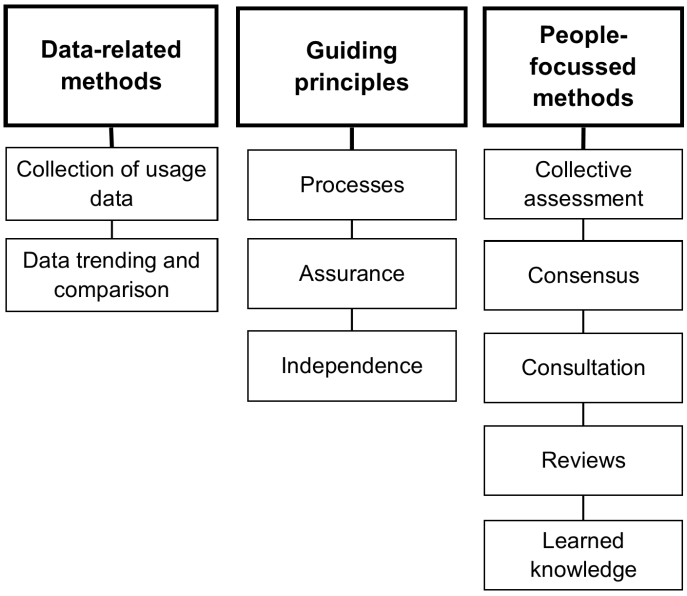


Figure 4.24: Survey Question 34 thematic analysis results.

Question 35 - "Is assurance activity carried out to validate the effectiveness of your structural integrity risk analysis?" The modal response was "yes" (71.9% (23/32) $\pm$ 15.6% (95% confidence interval)), and there were no patterns identifiable within respondent categories. Risk decision-makers all responded "yes" (68.8% (11/16) $\pm$ 22.7% (95% confidence interval)) or "unsure" (31% (5/16) $\pm$ 22.7% (95% confidence interval)) . Risk analysts were the only role to provide "no" responses (42.9% (3/7) $\pm$ 24.3% (95% confidence interval)), although the modal response for this role group was still "yes". Interview participants provided a richer perspective on these responses:

– Participant 58 (aviation, operator/end user, risk decision-maker) - *"Sometimes. . . I think it depends on the issue, I think if it's a really significant issues that's still on the radar, then, that's been mitigated but hasn't maybe been mitigated down to a level where it's considered to be low enough, then yes. But if it's something that you'd go 'right that's done' put to bed, then unless the issue re-rose it's head 10 years later you wouldn't look at it again."*

– Participant 61 (aviation, operator/end user, risk analyst) - *"...No because it's not formal assurance, other than the independent safety advisor, when*

people came in to so called assure us, they were looking at process compliance rather than reviewing our actual risk analysis, the decision we came to. So we could've been talking complete cobblers but as long as we'd followed the process it would've been fine, so I'm going to say no."

These responses suggest that assurance may be provided, but not necessarily to the depth of assuring the risk analyses, particularly if they are issues that are completed.

Question 35a - "Who conducts assurance activity?" *(Only "Yes" responses from Question 35.)* The modal response was "internal audit (internal to organisation)", but the significance of "external audit (external to organisation, but within same industry)" being a close second followed by "self audit" suggests that for risk analysis assurance to be effective, these analyses should be communicable within the same industry.

Question 35b - "Do you believe the assurance activity is effective?" *(Only "Yes" responses from Question 35.)* The overwhelming response was "yes". Only $n = 1$ "no" response was given, which was from a risk decision-maker in the operator/end user category. Interview responses provided further insight that suggested there are degrees to the strength of the "yes" answers provided:

– Participant 58 (aviation, operator/end user, risk decision-maker) - *"...More on the yes, because we do go over things at the [hazard review meeting], things like we have those meetings where we go back over bits and pieces in the [structural integrity stakeholder meeting] etc. so in terms of assurance, yes a lot of work is done to ensure that decisions aren't just made and put to bed. But it's only just on the yes, because for example if I made a decision on something and hadn't escalated it high enough to reach those meetings, it could easily not get looked at ever again."*

Question 36 - "What changes might be made to improve your organisation's structural integrity risk management?" The main themes regard changes that

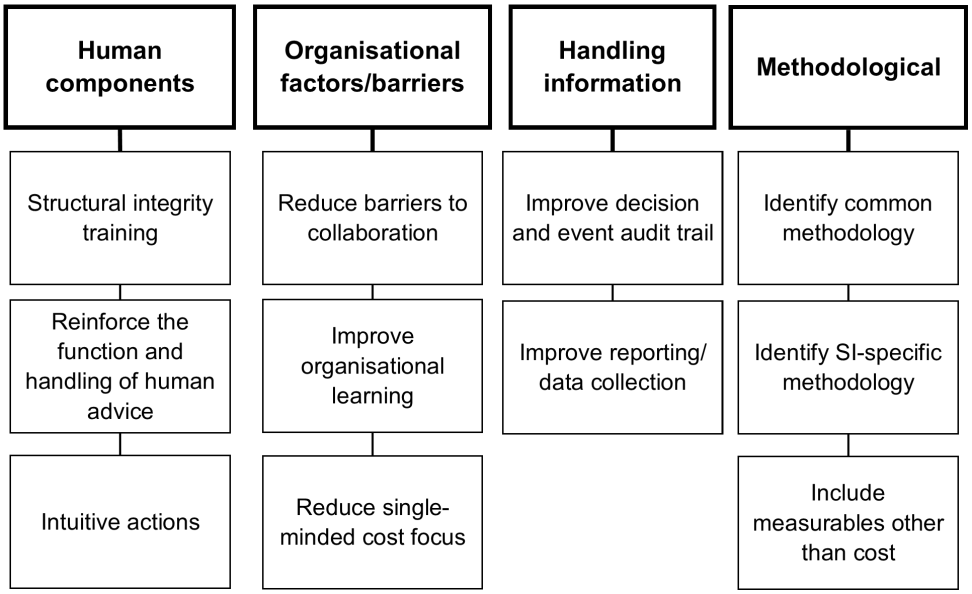


Figure 4.25: Survey Question 36 thematic analysis results.

target the human components in the system, organisational factors, the methods for handling fault and risk information, and enhancements to methodological aspects. The latter theme in particular highlights the desire for a common methodology in order to bridge organisations, equipment types and approaches to risk assessment, for example between subjective and non-subjective risk treatment to support non-subject matter experts. Figure 4.25 details the sub-themes.

Survey section 4 summary Investigating the management aspects of structural integrity risks highlighted that decision-making responsibility for structural integrity risk is, by a slight majority, collaborative (Question 29). Numerous institutional-enablers (procedures, stakeholders, recording requirements) lead participants to believe their treatment and ownership of risk was transparent (Question 30). However, rare events are only considered “to some extent”, with most participants not able to identify a formalised mechanism for identifying rare events (Question 32). Despite this, structural health monitoring is planned or in-use with a majority of participants (Question 33), and other data-related methods, principles or socially-based methods provide survey participants with confidence in their analyses (Question 34). Assurance activity was apparent among participants, although the resolution of the activity was not identifiable

(Question 35). Participants felt that improvements in the management of structural integrity risk could be found in supporting human actions in the handling of risk, overcoming organisational barriers, improving data handling, and the creation of a specific methodology (Question 36).

4.2 Phase 2 - case study results

To complement the participant-reflective survey data, empirical evidence was sought from appropriate real-world engineering case studies, selected using the criteria outlined in Section 3.3.1. Three cases provided sufficient perspective to explore the extent to which judgement influences the handling of engineering integrity risks. The first two are accessible in the public domain, the third required privileged access to the information owners (and their permission) within the UK Ministry of Defence and was selected to highlight typical structural integrity risk practices under operational conditions within the study's primary target domain. These cases were:

- The Challenger Space Shuttle launch accident, January 1986.
- The Lamington Railway Viaduct support collapse, December 2015.
- A UK Ministry of Defence example of an aviation engineering request for structural maintenance deferral, May 2017.

The situations are briefly summarised and one particular facet of each case is explored in detail using a template analysis to highlight the use of unsupported judgement under time pressure (Aven, 2018b). Table 3.1 lists the codes applied to the cases, and where a relevant line of text identifies with one of the codes, it is labelled in square brackets, for example:

> this line infers that perceptions about the risk were made [PERC].

4.2.1 Case study 1 - The Challenger Space Shuttle launch accident, January 1986

The Space Shuttle Challenger (mission number STS (Shuttle Transport System) 51-L) experienced a mid-launch complete vehicle destruction in January 1986 and the accident has been well documented (Rogers (1986), Vaughan (1996), Dekker (2019, p234)). The O-ring seals, which were designed to contain hot propellant gasses at the joints between sections of the solid rocket booster (SRB) had failed. The accident was preventable, and the post-incident investigations uncovered a series of conflicts between engineering and management personnel in the run up to a collective decision to launch, despite initial recommendations of 'no launch' and sustained dissent from specialist engineering personnel. In the wake of the accident, a Presidential Commission was convened (Rogers, 1986) to investigate the accident and all staff involved were subject to scrutiny by both the commission (under oath) and the public owing to media coverage (Vaughan, 1996, p388).

The case illustrates a major technical project that operated a risk management system and was subjected to operational stress factors for the Challenger launch scheduled for 28 January 1986. Although there is substantial discussion on this case that identifies organisational safety failures back to the design choice of the Shuttle vehicle, such as drift (Dekker, 2019, p234), the application of judgement unsupported by analytical evidence is widespread in the hours prior to the launch of STS 51-L. The rationale that led to the initial 'no launch' recommendation is studied here to highlight how the formal inclusion of judgement may have benefited the risk management process.

4.2.1.1 Limitations of the sources used

The secondary sources used provide contrasting narrative accounts of the decision to launch the Challenger shuttle. The Presidential Commission Rogers Report is criticised for being too focused, misunderstanding NASA procedures in decision-making, and not including some personnel's testimony in their consideration (Vaughan, 1996, p59), suggesting it is narrow in its conclusions.

But it provides timely and personal testimony from the protagonists involved in the flight safety decision process on 27 and 28 January 1986, which is the prime consideration in studying how far unsupported thinking contributed to the Challenger accident. To balance the use of the commission report, Vaughan (1996) provides an alternate perspective and source of wider testimony not included in Rogers (1986).

4.2.1.2 Narrative of events

Following Challenger's planned launch on 27 January 1986 being deferred owing to strong winds at the launch site, the launch was re-planned for 09:38 on 28 January. The forecast temperatures were colder than any previous launch. SRB manufacturing engineers at Morton Thiokol (MTI) and Marshall (responsible for maintaining oversight of the SRB contract among others) raised their initial concerns of the temperature effects on the O-ring seals between the joints of the rocket segments shortly after the decision to defer the mission 24 hours (Rogers (1986, p87) and Vaughan (1996, p286))[6]. These initial concerns triggered activity by both MTI and Marshall to collate the available data regarding temperature effects on SRB joint sealing, which was then discussed during two teleconferences between MTI, Marshall and NASA the night before the launch. During the second and more widely attended of these teleconferences, MTI initially presented a formal 'no launch' recommendation. Marshall rebutted the rationale on the grounds of inconsistencies between the data they presented and a history of behaviour by MTI who had not rejected launches given similar forecast temperatures[7] to those predicted on 28 January 1986 (Vaughan, 1996, p155, p308-310). Consequently, Thiokol reviewed their rationale and returned to the conference half an hour later with the conclusion that despite concerns over the cold effects on the O-ring seals,

[6]MTI's design of the SRB required the rocket to be manufactured as several discrete sections at their plant in Utah, USA in order to enable transport by rail to the launch site in Florida, USA. This logistical constraint meant that the SRB could not be manufactured as a solid single (monolithic) section. A monolithic design was simpler by comparison, which had been offered as a solution by competing contractors when the contract was awarded (Vaughan, 1996, p425).

[7]Mission STS 51-C was scheduled for launch on 23 January 1985, but was delayed by the Mission Management Team owing to the below freezing temperatures. No concerns had been raised by MTI or Marshall regarding the effect of cold temperatures on propellant seal integrity.

the data was inconclusive. MTI subsequently signed their telefax recommendation for the launch of STS 51-L on 28 January 1986. Challenger launched at 11:38 on 28 January 1986 at an ambient launch pad temperature of 36 °F (Vaughan, 1996, p7), exploding 73 seconds later in a fireball and with total break-up of the vehicle. The seals were found to have failed to perform their role during the first seconds of launch.

4.2.1.3 The initial 'no launch' rationale

The initial concerns raised about the cold by MTI and Marshall personnel infer an holistic, deliberative consideration regarding the primary cue (the forecast ambient temperature at the launch site) and the existing O-ring sealing risk. The O-ring seals had been subject to launch constraints and individual component testing after erosion and blow-by of the O-ring seals had been spotted following previous missions[8]. However, a normalisation of deviance within the team meant they had learned to accept some erosion of the O-ring seal during launch (Vaughan, 1996, p148-195). The fact that forecast temperatures prompted individuals to behave differently and with urgency for STS 51-L indicates that background assumptions (that were perhaps based upon on the *normal* conditions expected for a launch) had become challenged. The uncertainty regarding how the Shuttle's SRB seals would perform as a system, given the unprecedented cold, influenced multiple personnel to consider that the risk was high given their knowledge of previous instances of erosion and blow-by ([PERC], [UNHI] and [RHI]):

- Marshall's Larry Wear, who was the initial trigger for consideration of the risk from cold temperatures, recalled the effect of low temperature on STS 51-C in January 1985 and asked Boyd Brinton (MTI), to call Thiokol's Utah plant and find out if there were any concerns (Vaughan, 1996, p286).
- Team manager for the Solid Rocket Motor (SRM) project, Robert Ebeling

[8]STS-2 was the first indicator of erosion, heat effects were found on two rings on STS 6 (Vaughan, 1996, p149), and then in 1985 seven out of nine shuttles launched experienced erosion and/or blow-by, with the worst arising yet seen when STS 51-C's blow-by reaching the secondary O-ring (Vaughan, 1996, p153)

(MTI), responded to the request from Brinton by calling a meeting of subject matter experts to review his initial temperature concerns: *"The meeting lasted one hour, but the conclusion of that meeting was Engineering...were very adamant about their concerns on this lower temperature, because we were way below our data base and we were way below what we were qualified for."* (Rogers, 1986, p87)

- Allen McDonald (MTI) was at the Kennedy Space Centre: *"I took that data [the forecast temperature data] and called back to the plant and sent it to Bob Ebeling and relayed that to him, and told him he ought to use this temperature data for his predictions, but I thought this was very serious and to make sure that he had the vice president, engineering, involved in this and all of his people; that I wanted them to put together some calculations and a presentation of material."* (Rogers, 1986, p87)
- Other Marshall employees also expressed qualitative concerns about the effects of the anticipated temperatures on the performance of the O-ring seal:

  > *"Schell and Riehl concluded that the O-ring would be all right at 25 °F because when the ignition pressure hit it and jammed it into the gap, it would seal. Schell said: "It would have been a little harder, there's no question about that. I mean, there's data all over the world to show that it would have been a little less resilient and a little harder, but at those pressures, it would have sealed."* (Vaughan, 1996, p290)

In summary, uncertainty regarding the consequences of the weather compounded existing primary and secondary O-ring resilience concerns given the launch dynamics of the SRB joints, the ignition pressure and the time taken for the O-ring to form an adequate gas seal (Vaughan, 1996, p290). There was uncertainty owing to the lack of scientific data regarding cold effects on SRB joint sealing capability in full scale equipment configuration, and a conflict in the conclusions

drawn from component test results[9] and historical launch data[10]. Despite Thiokol's worries about the cold, no 'Launch Commit Criteria'[11] had been established for the O-ring temperatures after STS 51-C in January 1985 because "it was nobody's expectation we would ever experience any cold weather to that degree before we had a chance to fix it again" (Roger Boisjoly, MTI) (Vaughan, 1996, p308). Consequently, the only full scale vehicle temperature-related data of the SRB performance was perceived to show no correlation between temperature and the O-ring erosion and blow-by (see footnote 10)[PERC].

MTI were unable to support their 'no launch' argument with any analytical evidence because there was none for the whole SRB system at the forecast temperatures. MTI used judgement without analytic support to argue that cold temperatures had resulted in O-ring erosion on previous flights. However, because available O-ring performance data was apparently uncorrelated (footnote 10) Marshall and NASA were able to construct a convincing counter-argument. Notably, the influence of group dynamics on the decision made, is discounted by Vaughan (1996, p404). Regardless of the counter-argument, MTI engineers still expressed their fearful perception of launching outside of their experience base:

- MTI's Roger Boisjoly stressed that "in launching below the data base they were moving away from goodness". [PERC] (Vaughan, 1996, p317)
- MTI's Jack Kapp: "Most of the concerns that we had presented were qualitative in nature. At that particular time we had a very difficult time having enough engineering data to quantify the effects that we had been talking about. A lot of it was based on "engineering feel"." [PERC] (Vaughan, 1996, p308)

[9]The team had data on subscale component tests of resiliency (prevention of blow-by) to temperatures down to 30 °F, had conducted sufficient erosion resistance tests to conclude that the seal would still perform beyond the worst case seen from a live launch, and had data on how the O-ring seal material hardened with reducing temperature.

[10]STS 61-A launched in October 1985 and experienced blow-by and erosion at a calculated O-ring temperature of 75 °F. STS 51-C launched in January 1985 with a calculated O-ring temperature of 53 °F and exhibited the most severe O-ring erosion and blow-by from live mission launches (Vaughan, 1996, p153).

[11]Launch Commit Criteria were metrics that provided pre-determined thresholds and limitations beyond which a launch could not proceed.

Collective testimony coupled with post-event review of the information available to decision-makers shows the use of unsupported judgement regarding risk of cold temperatures to the Challenger vehicle. Without sufficient coherent analysis to substantiate the claim of safety, unsupported judgements were formed about the adequacy of the available technical data and the behaviour of MTI. These were subsequently found to be fallacious, with disastrous consequences.

4.2.2 Case study 2 - The Lamington Railway Viaduct support collapse, December 2015

Lamington railway viaduct partially collapsed following erosion of the river bed beneath the structural supports of the viaduct. Related risk information is traced back nearly 30 years to when another railway structure in the UK collapsed due to erosion of the riverbed beneath supporting structure, highlighting organisational failings to appropriately handle a known risk. The Lamington viaduct incident report was published on 14 Nov 2016, taking approximately 11 months to investigate, conclude and release the findings.

4.2.2.1 Limitations of the sources used

The incident report seeks to improve railway safety without establishing blame or liability (Rail Accident Investigation Branch, 2016, p3). Accordingly, the textual summaries of the actions can be considered factual as far as was investigated at the time. However, as a single source from one organisation's perspective, investigators may have not had access to particular evidence, the report may suffer individual bias (on the part of the investigators), or they may have been subject to legal constraints at the time.

4.2.2.2 Narrative of events

On 31 December 2015, prolonged heavy rainfall and subsequent high flow in the River Clyde, Scotland, UK, created the conditions necessary for the erosion and

removal of river bed material under the base of the Lamington viaduct. The erosion phenomenon known as ‘scour’ caused subsidence of the viaduct support structure, which became noticeable by train drivers passing over the viaduct that morning as a ‘dip’ in the track. Attending track maintenance engineers initially diagnosed minor track deformation as the cause for the reported ‘dip’, but closed the line when a passing train created surprisingly large track movements. By this point, several trains had been permitted to travel over the viaduct. Subsequently, the maintenance team found damage to the central pier of the viaduct, which was the root cause of the track deformation. The trains that had passed over the damaged viaduct had been fortunate to escape derailment or to cause a more dramatic collapse of the viaduct. The incident was declared a “dangerous occurrence” (Rail Accident Investigation Branch, 2016, p26).

The viaduct had been subject to multiple formal assessments since 2005, consistently scoring as a ‘high risk’ structure by a contracted surveyor using a network-wide scour priority rating system. Annual underwater inspections provided routine mitigation activity and extreme weather precautionary measures provided reactive procedural mitigation (such as monitoring river levels and closing the line if a threshold waterline was reached), but these had fallen out of use owing to organisational changes. Planned scour reinforcement works to the viaduct were deferred from 2015 until mid-2016 owing to essential environmental approvals not being secured.

4.2.2.3 Judgement-based decision points

Two decision points highlight where judgement was applied under operational constraints that influenced the handling of the risk of viaduct collapse.

Scour reinforcement works Two contracted formal initial scour risk assessments (in 2005 and 2010) had classified Lamington as at high risk of scour. Despite knowing the formally assessed risk, management elected not to apply a full and timely risk intervention [NBPA]. Recommendations to conduct a follow-up detailed assessment and regular physical scour and flood damage

checks were only partially complied with. Mitigating annual underwater inspections only provided 'point-in-time' conditional monitoring of the viaduct support structure, and repair work that was called for took three years to install and was temporary in nature. Despite the actions selected, Lamington's listing on a schedule for scour protection works in 2012 indicates that the risk influencing factors were known about: "Lamington was prioritised due to its location on the West Coast Main Line and known shallow foundations" (Rail Accident Investigation Branch, 2016, p33).

This shows business and reputational concern for the importance of the route that the Lamington viaduct was situated on ("its location on the West Coast Main Line") and knowledge of the formal risk assessment findings ("known shallow foundations"). A third and more detailed scour risk assessment was conducted in 2013, again scoring Lamington as high risk. The description of events infers that the decision-makers have used the risk analysis as a constituent in a wider value-focussed decision-making process (Hansson and Aven, 2014), since rail authority personnel "requested authority to defer the scheme until the next financial year due to environmental permissions not being sought in time for the work to be completed during the low flow summer period" (Rail Accident Investigation Branch, 2016, p33) [NBPA].

The lack of timely action suggests the decision-makers did not perceive the risk (or uncertainty surrounding the formal assessment conclusions) from scour to be sufficiently high to cross their personal threshold for intervention ([RHI] and [UNHI]). Decision-makers may have been mistakenly optimistic from the annual underwater inspections and overly reliant upon assumptions regarding the adequacy of risk mitigation that was in-place. When scour reinforcement works were deferred in 2015, the stated reason was that there were "no new defects noted by the latest visual examination" (Rail Accident Investigation Branch, 2016, p33). This would be reliant on assumptions such as observable gradual deterioration preceding a total collapse but not deteriorating catastrophically between inspections [BGKN]. Other influences on this judgement may have been perception of the true risk to Lamington, assumptions hidden in some personal knowledge of the structure in place at Lamington, the behaviour of the River Clyde, or of the mechanics of scour

risk ([NBPA] and [BGKN]). Decision-maker judgements were made *in spite* of the formal analytical risk assessment and there is no evidence of the grounds for why robust intervention activity was so tardy.

The inability to substantiate structural risks and prioritise mitigation activity highlights the difficulty that decision-makers have in communicating the urgency of risk intervention requirements on structure to resource managers and budget approvers. Meaningful intervention was only initiated once a clear unsafe threshold was crossed - when damage was visible. Despite quantitative assessment scoring Lamington as a high priority, qualitative value-based judgements rationalised continuing deferment of intervention activity. This demonstrates unsupported use of judgement in spite of an available analytical risk assessment.

The on-call track maintenance engineers close railway On arrival at Lamington to assess the reported track fault, the maintenance team's combined experience initially misdiagnosed minor 'dips' in the line. This deliberative conclusion led to the exposure of three trains to a structurally-defective viaduct. The decision may have been reached because the maintenance team believed (or assumed) the track dips were isolated and the most likely root cause from previous experience or historical evidence: "the vast majority of track faults are directly related to track condition" (Rail Accident Investigation Branch, 2016, p35) ([LA], [NBPA] or [BGKN]). The maintenance team's lack of knowledge regarding scour risk and the specific threat posed to Lamington (Rail Accident Investigation Branch, 2016, p20) was also a missing cue. The decision was unsupported by any formal analysis, made dynamically and on the basis of underlying assumptions being true [BGKN], such as the fault being track related only, the supporting structure being unquestionably sound and with no other external influencing factors. The grounds for these assumptions were then found to be false once the maintenance team identified the structural damage to the viaduct. Their reaction was to close the

line immediately, in line with procedures[12]. The team were not bridge specialists (having requested the attendance of a bridge examiner and being unaware of scour risk) but perceived the risk to be high, fearful or unacceptable once they had visibly identified structural deterioration [PERC].

The non-structural-specialist track maintenance engineers were required to make a timely decision on an unfamiliar risky prospect, which was known about within the organisation's risk management framework but not by the track engineers. Their decision did not make use of existing formal risk analyses that had been carried-out on Lamington Viaduct, and required the use of judgement to escalate the situation for wider organisational consideration.

4.2.3 Case study 3 - UK military aircraft routine structural damage

At the time of the case study, the UK military aviation regulations required risks to life to be managed within a risk management system, with nominated duty holder personnel being accountable (Military Aviation Authority, 2015). Authorised engineering management personnel are also permitted to defer maintenance provided it is "considered justifiable and safe" to do so (Military Aviation Authority, 2016, p8)[13]. Deferring maintenance is considered routine since spares, operational considerations, specialist equipment, or expertise are not always available at the point of need for a given maintenance task. Case 3 is one such routine request sent to an aircraft airworthiness management team from front-line engineers operating the aircraft. Although it does not end in an accident, it highlights the routine use of judgement when handling structural risks in operational situations.

[12]It is noteworthy that railway signalling and control room staff would not have been permitted to allow trains to be exposed to *suspected* structurally damaged equipment (Rail Accident Investigation Branch, 2016, p35), however control room staff were also unaware that Lamington viaduct had been assessed as being at high risk of scour and therefore susceptible to structural damage. Although this suggests an information and communication breakdown, it also demonstrates the conservative intervention required in the face of insufficient structural risk information (Aven, 2016b)

[13]The latest revisions of these regulatory documents (Military Aviation Authority (2018a), Military Aviation Authority (2020b)) remain consistent with these requirements.

4.2.3.1 Limitations of the source

The information was received directly from the engineering staff responsible for the structural airworthiness considerations of the particular military aircraft. Much of the detail is anonymised to protect individuals and teams from scrutiny and to prevent any breach of confidential information regarding the sustainment of military equipment. Although this restricts the level of clarity experienced by the reader, there is sufficient information to identify the use of unsupported judgement in daily aircraft engineering risk decision-making.

4.2.3.2 Narrative of events

On 3 May 2017, a request was submitted by the operating Squadron to the aircraft airworthiness authority and manufacturer to defer carrying-out a full repair as required by the aircraft maintenance manual. A structural repair was necessary after accidental tooling damage was found on a shaped component that provides the structural connection support between the wing box and the outer wings of the aircraft. The component is known by the manufacturers to be susceptible to fatigue and the damage was spotted during a periodic non-destructive testing (NDT) inspection that preventatively searched for fatigue damage. The operating Squadron were only able to carry-out a partial repair as they were deficient in the necessary tooling to complete the repair at their location. However, as no damage limits for the component were provided in maintenance manuals, a request for airworthiness advice and legal maintenance data was dispatched. The Squadron had no legally approved maintenance data that a partial repair would be sufficient not to cause a structural integrity hazard to the safe operation of the aircraft. A NDT inspection of the repaired area was carried-out as mitigation for the absence of a full repair. The Squadron requested:

“Airworthiness Advice on continued unrestricted Operations... Tooling impact damage to R/H upper [component name omitted for confidentiality]. Pre- blended

damage[14] *dimensions as follows [dimensions omitted for brevity]. Damaged areas blended iaw (in accordance with) [structural repair manual reference omitted for confidentiality] to a ratio of 10-1. NDT [reference omitted] carried out with no fault indications. Roto peening*[15] *not carried out due to lack of tooling in theatre*[16]*. Post blend dimensions as follows [dimensions omitted for brevity].”*

While awaiting a response, the Squadron deferred a full repair and continued operating the aircraft unrestricted. A month later (5 June 2017), approval for the partial repair was received from the manufacturer, via the airworthiness management team. In the interim, the acceptability of the risk to the aircraft had been based on the judgement of the front-line maintenance decision-maker without analytical or deterministic confirmation.

4.2.3.3 Judgement to defer maintenance

Deficiencies in equipment, spares or information to complete engineering activity are frequent in military aircraft maintenance. While designers foresee problems as far as possible and maintainers plan equipment usage and manpower schedules as thoroughly as possible, aircraft operations inevitably generate unexpected damage as a result of human and environmental influences. Deferring maintenance at the point of need allows maintenance managers to use their judgement to balance safety and equipment availability, catering for real-world variations in situation while awaiting manufacturer advice. However, without access to the tools or cognitive ability to generate analytical structural integrity risk assessments, maintainers must proceed non-analytically.

Aircraft maintenance is typically documented scrupulously, recording references to procedures, policies and information used, providing “an audit trail of the work to enable quality assurance, data exploitation and investigations” (Military

[14] Blending is a repair technique that involves the abrasive rubbing and polishing of metal damage in order to smooth out damage such as dents, gouges or scores. Sharp edges create ‘stress raisers’ in materials under load, and can be initiators of fatigue or overload damage mechanisms.

[15] Roto peening is a specialist maintenance activity that imparts a residual compressive load into the surface of metal structure in order to provide tensile stress relief. The process improves resilience to fatigue and some corrosion damage processes.

[16] ‘Theatre’ refers to a deployed military operational location overseas.

Aviation Authority, 2020c, p1). The omission of references to quantifiable data, relevant historical events, or analytical assessment supporting the deferment request in Section 4.2.3.2 is conspicuous in its absence. If analytical tools had been used, they would almost certainly have been referenced in this maintenance paperwork entry. This suggests that a judgement on the acceptability of the structural risk was made by the engineering manager, given their personal state of knowledge regarding factors such as: 1) the assumed expected use of the aircraft, 2) the available knowledge and experience of the damage characteristics, 3) that the NDT inspection was suitable and accurate, 4) no other assumed factors would interact with the damage, 5) the impact on remaining structure if the damaged component was to fail. The factors that influence damage propagation are numerous and situationally specific. Decision choices also vary between decision-makers - another engineer faced with the same situation but with a different knowledge base may elect to ground the aircraft while awaiting specialist advice. Descriptive decision analysis provides insight into the motivational and cognitive biases that may cause this (see Section 2.3.4). For example grounding an aircraft requires just as strong an argument as a decision to accept a fault and continue flying because grounding aircraft impacts operational availability, may cause reputational harm (depending on the service the aircraft is providing), and commits organisational resources to rectifying the fault. In this case, the judgement was that the risk posed by the partial repair did not exceed any personal thresholds for intervention. However, if the risk had been deemed unacceptable, the considerations influencing the operational decision-maker may have been:

- Mitigation being considered inadequate, for example the NDT technique might provide insufficient assurance or be considered unreliable [UNTHI].
- Previous experience of this type of repair technique might bias the judgement toward the success or failure of the experience base. This might be considered as a learned association [LA] or an assumption that the repairs are sufficiently similar to draw an inference between the damage situations ([BGKN] and [NBPA]).

– Other personnel may offer up their own experiences, which the manager will weight given their knowledge of the advice-giver, relying on social factors such as reputation and trust [LA] .

Such influences are very personal and impact people to varying degrees. In the case above, the operational urgency and perception of the risk meant that the maintenance manager resorted to their judgement and perception about the risk [PERC], exercising their authority to continue operating while awaiting supportive information.

4.3 Survey and case study discussion

The purpose of Phase 2 was to capture an inductive description of how engineering practitioners treat structural integrity risks when operationally constrained by time, information and resource. In doing so, this provided a rigorous confirmation of the author's experiences regarding structural integrity risk management in aviation maintenance. The primary data collection via survey enabled individualistic perceptions to be collected in order to identify similarities and differences across organisational roles, while the analysis of case studies offered complementary evidence of industrial practice through retrospection.

The survey results were predominantly descriptive in nature, though differences between groups were assessed using statistical tests for significance. This offered insight into the differences between participant groups, although the statistical strength of the results was not high enough to identify anything other than large effects. This may have been a consequence of the number of participants that were able to be recruited for the survey.

Analysis of case study data was supported by codes extracted from Aven (2018b) that connected dual process theory of decision-making to risk analysis. The codes were used to categorise human perceptions used in risk analysis that was identifiable from the case study data.

Reviewing and combining the results of the survey and case studies, the two data sets coalesced around primary themes related to the first

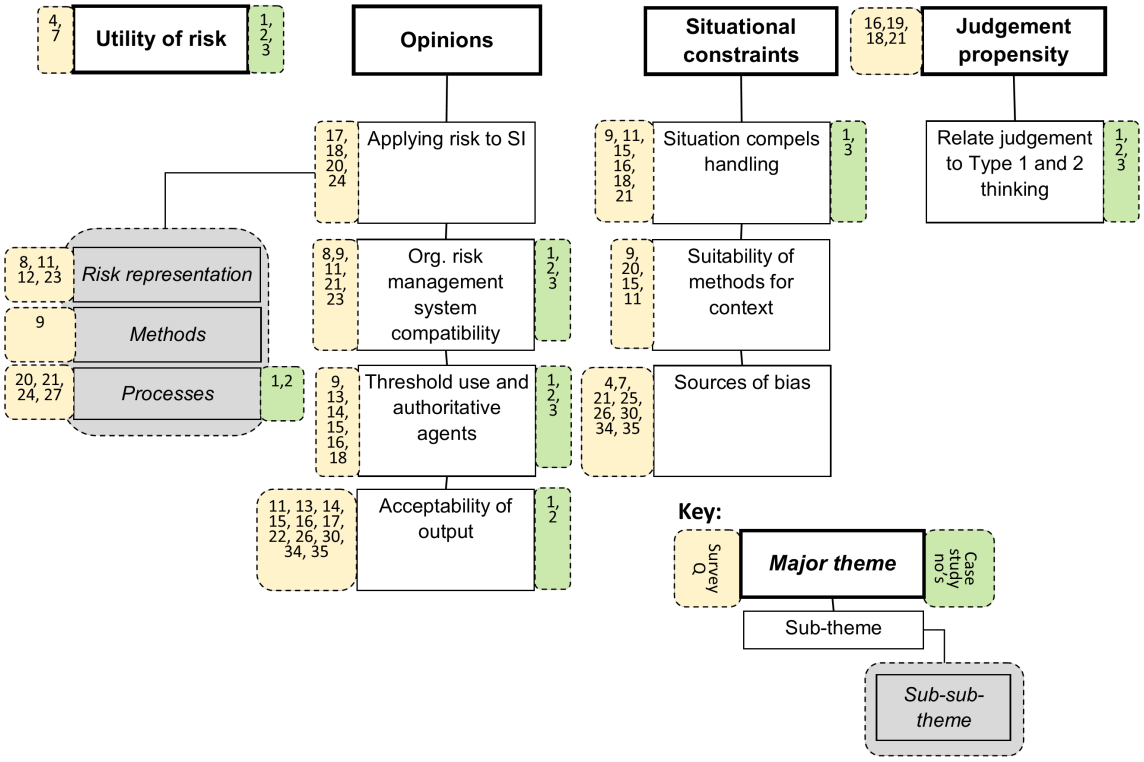


Figure 4.26: Initial data capture discussion thematic map showing the survey questions and case studies related to each theme.

research objective: *gain a detailed understanding of the beliefs and practices of real-world practitioner decision-makers deliberating structural risks in operationally constrained conditions*. Each survey question was considered individually for meaning and for triangulation with other survey questions and the observations made about risk judgement-use in the case studies (identified in Green et al. (2021)). Sub-themes were connected through related parent themes, and the resulting thematic map showing the four primary themes listed below is shown in Figure 4.26.

- Theme 1 - The utility of risk concepts.
- Theme 2 - Differing opinions.
- Theme 3 - Situational constraints.
- Theme 4 - Judgement propensity.

These four topic areas structure the discussion within this section and aid understanding relevant to the first research objective. The themes also framed deliberations during Phases 3 and 4 of the research, although are not referred to directly. This section is laid out as a discursive summary of the survey and case study results. Relevant internal references are flagged in rounded brackets, for example:

> "Organisation risk management policies are not always configured to handle structural integrity risks" (survey Question 8c, 8d, 9 and 23; case study 1 and 2).

A summary of structural integrity risk handling from the perspective of operational engineering practitioners (as informed by the survey results) is provided in order to frame the follow-on discussion about each theme identified.

4.3.1 Overview of structural integrity risk handling

Structural integrity practitioners indicated value for risk activities, although there were differing opinions about whether risk approaches were appropriate for structural integrity failure prediction. The risk analysis and decision processes followed were largely common despite nuanced differences in individual responses. Respondents also converged on a number of issues regarding risk as applied to structural integrity, discussion of which overlaps the main themes identified above.

4.3.1.1 Perspective on value

There is widespread agreement that organisational risk management policies consider structural integrity risk (survey Question 4) and that this is valued (7). However, a minority proportion of practitioners do not believe structural integrity risk is specifically considered (4). Survey respondents indicated that risk analysis deliverables and the process are valued for improving competence, understanding consequences, decision-making and meeting cultural, moral and policy obligations (7). The survey also identified consensus about the group considered the main benefactor of risk assessment outputs: operator/end users (31).

Value in maintenance risk management approaches was also inferred from all of the case studies. The Shuttle programme operated within an extensive risk framework managed by NASA that defined strict acceptance criteria. The formalised system emphasised the importance of empirically validated test evidence in order to inform decisions about risk (case study 1). The railways authority commissioned three separate formalised risk analyses of the viaduct, at expense and commitment to the railway authority (case study 2). The aviation maintenance case was within a risk management system that assigned accountable persons, responsible for the management of risks to life (case study 3). The aircraft fault was handled within the constraints of regulatory policy, and was ultimately considered by the decision-maker as sufficiently low a risk to permit continued operations. Although the outcomes were case-dependent, the value of risk processes to organisations is readily apparent.

4.3.1.2 Processes

Systematic approaches to handling structural faults as risks are apparent. Commonality existed across respondents in the approach to risk analysis (survey Question 6), following a process of: identify the risk; search information and assess the situation relevant to the information found; decide and act; follow-up on chosen actions. Some overlap was also identified with the decision processes (20): *identify* the issue; *assess* and classify the situation; *intervene*. However, these processes were extracted as higher level abstractions from all answers provided, obscuring the individuality of responses. For example, in survey Question 20, participant 2's (from the designer/equipment manufacturer group, risk decision-maker) process was identified as:

> *"Assess flight safety impact; analyse damage deterioration; assess structural integrity risk; intervene"*

As a comparison of similarly categorised respondents, Participant 50's (also from the designer/equipment manufacturer group, risk analyst) process was identified as:

> *"Urgency/severity assessment; hazard risk index assessment (aggregated with intervention activity); use policy definitions to determine hazard risk index"*

Both of these attend to the holistic process of *identify/assess/intervene*, although the discrete elements of the process are specific in nature.

There were specific and differing opinions about the management of structural integrity. From practitioner decision processes (Figure 4.18a) and the intended outcomes from structural integrity risk analysis (Figure 4.18b), fault management either included risk parameters or depended on non-risk/deterministic solutions. A subset made specific consideration of situational/operational factors. A common finding was that deliberations about risks are, for the most part, not isolated from decision-making, both examples above including value aspects: flight safety impact (Participant 2); urgency/severity (Participant 50). Both examples above also include a discrete risk management step: "assess structural integrity risk" (Participant 2); "hazard risk index assessment" (Participant 50). This may be a consequence of risk analysis and decision-making requirements converging under operational constraints.

Risk decision processes are not identifiable from the case studies. However, the approach taken by the decision actors was the same: a fault was manifest that was identified as a hazard to the safe operation of the equipment, and a decision was made regarding the extent of the intervention necessary.

4.3.1.3 Issues

Survey respondents reported their difficulties in applying risk management approaches to structural integrity faults, which converged on 4 areas. These overlap with the discussion section themes, but the results are briefly introduced to highlight practitioner perspectives.

Specific consideration of structural fault circumstances Despite widespread value for risk activity, specific appreciation of structural faults and its nature appears to be lacking. Participants reported difficulty generating likelihoods

(survey Question 11), had differing opinions about risk representation (8c, 8d, 23) compared to organisational requirements (8, 9), and there were calls for a structural integrity-specific methodology (36).

Consistency of risk assessment and intervention Differing opinions about fault intervention approaches were highlighted by differing decision processes (20, case study 1) and perceptions regarding acceptable risk threshold authorities (9).

Situational influences upon action Qualitative responses highlighted the influence that situational factors had on practitioner risk perspectives: time available (19, 20, 21, case studies 1 and 3), resources accessible for intervention (19, 21), information (9d, 19, 20, 21 and 26, case studies 1 and 3), access to advice (19, 20, 21), suitability of processing methods (9d, 9e, 19) or urgency (19, 20, 27a, case studies 1 and 3).

Difficulty in the application of judgement during risk assessment

Judgement use is widespread, indicated by qualitative[17] and quantitative (16, 21) survey results. The reliance on judgement (21) may be a causal factor in the difficulties faced generating likelihoods (11), handling assumptions (13) and uncertainties (14). The use of judgement was not always communicated (16a). These indications are a possible sign of conflict with organisational requirements for risk, where participants may have difficulty generating numerical risk representations required by organisations (8, 9). Case study 1 evidenced specialist engineers who were unable to effectively communicate and justify their deliberative judgements about cold temperature effects on the O-ring seals to a level of precision sufficient enough to be convincing. Case study 2 identified a situation where assumptions regarding the continued relevance of risk mitigations were found to be false once the viaduct began to collapse.

[17] Qualitative survey themes identified: *use of judgement is widespread* (survey Question 6); *individualistic perspectives and subjectivity* (8b); *judgement/descriptive/qualitative* (9a); *human influence on threshold for action/intervention* (9e); *individualistic aspects* (18); *individual judgement threshold* (19); *visualising or determining sequence of events* (20); *it (judgement) is always required* (21a); *individual judgement* (22a1); *individualistic decision-maker qualities* (26a).

4.3.2 Theme 1 - risk concept usefulness in practical settings

Despite difficulties in establishing a meaningful risk representation, considering structural integrity hazards from a risk perspective is useful to practitioners. Survey Questions 4 and 7 were situated at the start of the survey to prevent later probes (in to the use of particular risk features (11-16), judgement use (21) assessment accuracy (8a) and justifiability (22)) influencing reflection on the value of structural integrity risk activities within respondent organisations.

Responses to Questions 4 and 7 highlight the positive perception of the respondents towards structural integrity risk as is applied within their organisation. However, Question 4 was worded in a way to be deliberately unspecific and does not identify the degree to which organisation policies consider structural integrity risk. For example, an organisation that requires risk processes to be applied for all aspects of engineering maintenance (inclusive of mechanical systems, communications, avionics etc.), but does not account for the individuality of structural faults, may be viewed by the respondent as not considerate of structural integrity specifically. This was left to the participants to consider to gain a view of individual perspectives, not assess organisation policies against an arbitrary measure of quality. Question 7 responses indicated overwhelming agreement that structural integrity risk is a valued activity, although the nature of “value” in this context is not clear. Whether value is meant as a principle, value of a usable output, or value in the ability to apply risk practice (such as through hazard identification), the distinction is not necessary to understand that risk activities are valued by practitioners.

This was observable in the three case studies. All of the examples were taken from contexts inclusive of an organisational risk management policy, and showed that the risk management activites were valued, as discussed in Section 4.3.1.1.

4.3.3 Theme 2 - Differing opinions

Although objective risk analyses enable a platform for unambiguous measures of uncertainty about hazards, the survey results and case studies illustrate the

subjectivity inherent to structural integrity risk assessment. From methodological perspectives regarding probabilistic interpretations suitable to a given situation, to organisational perspectives on the suitability of risk frameworks to a particular situation, to individualistic perspectives on the benefit of risk concepts to a given issue. Real-world operating environments see the convergence of humans, technical equipment and forces in the natural world, creating a sociotechnically complex network of influences and possibilities. Differing perspectives on how structural damage may propagate are dependent on knowledge about past and future use, understanding of the operating environment, and specific structural mechanics knowledge. These factors that influence opinion were identifiable from the results:

- The application of risk principles to structural integrity,
- Organisational risk management system compatibility,
- Intervention threshold use and authorities,
- The acceptability of risk assessment outputs.

4.3.3.1 The application of risk principles to structural integrity

Despite widespread agreement about the value of structural integrity risk assessment (survey Question 7) and a majority of practitioners having confidence in the risk analysis methods they apply for operational risks (17), some participants did not believe that risk approaches can be used to handle structural integrity issues. Participant 43 (aviation, no category provided, advisor role) commented:

> *"The questions posed* [in the survey] *seem to display an underlying belief that risk can be quantified and in that sense may give rise to biased responses. Although structural risk as considered in this survey is unquantifiable applied Engineering Judgement should be strongly founded upon a detailed understanding of structures and materials through all phases from the determination of*

> *structural properties through design, construction, test, analysis and qualification. Once this level of understanding is matched to a realistic appreciation of how aircraft are maintained and operated the degree of Engineering Judgement that results can be relied upon."*(18)

Similarly, Participant 12 (aviation, designer/equipment manufacturer, risk analyst) commented:

> *"In structures we do not have analysis methods to calculate risk, the analysis has been answered as judgemental methods, supported by static and fatigue calculations."*(18)

These perspectives are framed as decision processes that lie on a spectrum between non-risk process focus and risk process focus (20). The difference in perception is significant when reviewing how risk is used for decision-making: are decisions in organisations risk-informed or risk-based (24)? There was a difference within the survey results between risk decision-makers and non-risk decision-makers (test for two proportions, Fisher's exact method, $Z = -2.28, p = 0.085$), which suggests a difference of opinion in the suitability of structural integrity risk assessment outputs for supporting decision-making. However, as this is not significant, this finding cannot be generalised without a larger sample.

Risk representation Survey respondents indicated that in some cases, only an incomplete characterisation of risk can be achieved (survey Questions 11 and 12). Evaluative differences regarding the risk representation necessary for the assessment and management of structural integrity risk, qualitative or quantitative, were also identified (8c, 8d). Polarised individualistic preferences highlighted that there was no clear majority amongst practitioners for either verbal or numerical forms (23), risk decision-makers showed a preference towards "verbal" representations, although this was not significantly different to non-risk decision-makers at the 5% level (test of difference between two proportions, Fisher's exact method, $Z = 2.12, p = 0.094$) (23). Larger sampling would confirm the extent that individual roles mediate perceptions about how to use structural

integrity risk analyses in decision-making. Some practitioner preferences (8c, 8d, 23) are in conflict with organisational requirements (8). For example, where an organisation necessitates quantitative forms but the individual has a preference for verbal forms or does not believe a quantitative form is necessary for managing structural risk. This is discussed in Section 4.3.3.2 below.

In addition, differences between survey groups and categories suggest variation in the ability to generate risk parameters for structural faults. Operator/end users, for example, were more likely to disagree that a clear likelihood could be generated (11), although the result was not significant ($W = 101.5, p = 0.087$).

Methods Unsurprisingly, there were differences in the methods used by practitioners handling structural integrity risk[18] (survey Question 9d). Operator/end users showed a greater preference towards qualitative methods, and designer/equipment manufacturers showed a preference toward classical statistical and stochastic approaches, although both groups also used other methods. This is perhaps suggestive that organisation function provides some predictive power over method selection, although the intent of this question was only to provide a description of the common methods used. These observations are expected, since designer/equipment manufacturers are more likely to have access to privileged data that enables statistical methods to be used.

Process Different decision approaches are used in handling structural faults and failures (survey Question 20). One explanation for differing beliefs over the treatment of faults as deterministic or using risk approaches may be that each individual is operating within their own paradigm of decision-making. Considering decisions on a spectrum of time and impact, decisions can be grouped: strategic (planning); operational (planning); instantaneous (execution); and emergency (execution) decisions (Yang and Haugen, 2015). This frames the survey participant decision-making processes: how risks and interventions are decided upon are conditioned by individual perspectives about the timescale and impact of structural

[18] There is insufficient statistical power to generalise the findings for Question 9d: design/equipment manufacturer $n = 4$, operator/end user $n = 8$, and regulator/policy maker $n = 7$.

risks. Patankar and Taylor (2004, p27) made a similar observation of the differing perspectives of safety and the resulting influence on the risk narrative across aviation actors.

Another explanation may be that individual approaches to risk are dependent upon considerations about risk assessments being non-compensatory (the deciding factor, see Payne et al. (1993, p29)). Risk decision-makers preferences for risk-informed decision-making (24) suggests that this group includes risk assessment as a compensatory cue, among other cues. Although the result to Question 24 was not significant, case study 1 was also indicative of this effect. The actions recorded during the discussion about the Shuttle launch regarded disagreement over the risk to the O-ring seal integrity given the forecast temperatures. The existing risk management processes that were empirically dependent did not pacify the discomfort felt by a number of MTI engineers about the decision to launch:

> MTI's Jack Kapp: *"Most of the concerns that we had presented were qualitative in nature. At that particular time we had a very difficult time having enough engineering data to quantify the effects that we had been talking about. A lot of it was based on "engineering feel". "*(Vaughan, 1996, p308)

This suggests the ambient temperature at launch was a non-compensatory cue for the MTI engineers. In the viaduct case (case study 2), the argument made for investment to reinforce the viaduct support structure did not attract priority action, despite being found by formal risk analysis as at high risk of scour. Although not explicitly documented, other factors may have contributed to decision-maker hesitancy to commit to action in both cases. The processes applied by the protagonists highlight the different decision approaches used, given the risk information available.

A final explanation for the differences in processes may be individual comprehension of structural integrity. The reasons given for decision-makers not understanding assumptions and uncertainty included individual competence (27), and respondents commented on difficulties experienced when personally

processing risk information using their judgement (21c). It is a reasonable assumption that not all maintenance decision-makers will be structural integrity experts, so interventions chosen to address structural failures and faults are a consequence of individual expertise.

4.3.3.2 Organisational risk management system compatibility

There are differences between the requirements of the organisational risk management system and individual approaches to risk (survey Question 8, 8c, 8d, 9, 11, 21b1, 21c, 23). Published policy or guidance material expects rational action by employees through conformance with recognised best practices. However, practitioners who necessarily deviate from procedures, perhaps because they are unable to comply, demonstrates localised rationality. The results highlight a gap between how risk assessment is imagined to be carried-out, and how it is able to be carried-out, a known effect in safety science (Dekker, 2019, p50). Some respondents indicate preferences and beliefs about structural integrity risk (8c, 8d, 11, 21b1, 21c and 23) that conflict with the form of risk assessment that their organisations require (8 and 9). For example, where participants answered "verbal" for Question 8[19], but "numerical" for Question 23[20], there may be a conflict between the representation an individual is most comfortable with for decision-making and that required by their organisation for risk management. Table 4.2 identifies the proportion of responses in each question that are in tension with answers given to Question 23.

Conflicts within the survey may be the result of respondents considering their answers at differing levels of abstraction. They may also be the result of a cognitive dissonance regarding the concept of risk analysis as applied to structural integrity, or a belief about wasted resource or time being committed to an unachievable analysis form. Participant 50 (aviation, designer/equipment manufacturer, risk analyst) highlighted their processing difficulties experienced when using judgement

[19]Question 8 - "Within your organisation, are structural integrity risks that arise in-service described numerically or non-numerically / verbally / descriptively?"

[20]Question 23 - "Are you more comfortable with numerical or non-numerical/verbal/descriptive representations of risky decisions regarding in-service structural integrity?"

Question	Proportion in tension
8	14%
8c	50%
8d	47%
21b1	7%

Table 4.2: Proportion of participants that indicated differing preferences about risk representation, in comparison to organisational prescriptions about risk representation.

to support structural integrity risk decisions:

> *"For most structures risks difficult to assign a probability for each stage of a credible route to failure. Thus tend to use descriptive assignments in Company HRI [hazard risk index] matrix. The HRI process has been devised for engineering subject matter with failure rates etc. and so does not always lend itself to Structures."*

In the case studies, the risk decisions were made under situational constraints that were incompatible with the required organisational risk management processes, compelling decision-makers to use unaided judgement in their decision. The initial decision by the Lamington viaduct track maintenance team caused them to react with an assumption regarding track deformation without knowledge of the known scour risk (case study 2); and the aircraft Squadron engineering decision-maker made an unaided judgement that the partial repair would not lead to structural collapse (case study 3).

In the case of Challenger (case study 1), the decisions, justifications and testimonies indicate that the risk management process was unable to accommodate the holistic and unsupported judgement that prompted numerous key personnel to initially argue a 'no launch' rationale in spite of the technical data available. This observation is not just because the correct decision (in hindsight) had not been selected, but because the risk management system was not configured to give weight to unsupported assessments in operational circumstances (Hoffrage and Marewski, 2015, p148). Within NASA's existing

technical culture and processes, risk assessments followed scientific and rule-bound engineering standards, but Thiokol's argument conflicted these:

"...observational data, backed by an intuitive argument, were behind all engineering analysis. But subjective, intuitive arguments required lab work and tests before they were considered admissible evidence in FRR (Flight Readiness Review)." (Vaughan, 1996, p353-354)

In all cases, the general representations of risk perceived by individual engineers provided an actionable view on the risk situation but in cases 1 and 2, the qualitative assessments could not reduce the risk exposure sufficiently. Furthermore, the hazards and risks were known to the organisations in cases 1 and 2. They were not black swan events that surprised stakeholders once they emerged, but the uncertainty surrounding risk propagation created sufficient latitude for arguments both in favour and against intervention. Unable to argue convincingly with judgement-based risk representations to gain support for timely intervention, the risks were unable to be prevented from becoming manifest.

4.3.3.3 Threshold use and authority identification

Acceptable risk thresholds are promulgated by organisations to control risk ownership and manage the resources allocated to risk management (for example the military policy in Military Aviation Authority (2018b)), ultimately to minimise harm. In the survey results, a lack of consistency was apparent among practitioners regarding which organisation is responsible for setting risk-based thresholds (9), and the relevance of thresholds to structural risks was disputed. Participant 31 (aviation, operator/end user, risk decision-maker):

> *"'Clear' consequence and likelihood are extremely subjective, applied judgement and experience will often determine the threshold (as opposed to defined, measurable metrics)"*(18)

Of the 20 participants who stated they work to a threshold (9), $n = 12$ (63.2%)

did not agree their method generated a clear likelihood, $n = 18$ (90%) more often than not include uncertainties in their structural integrity risk analyses, $n = 17$ (85%) more often than not include assumptions in their analyses, $n = 20$ (100%) handle risks conservatively, and $n = 19$ (95%) believe to at least some extent that human subjective judgement influences structural integrity risk analyses. Practitioners may be comparing structural risks to thresholds with an unknown level of precision, obscured by the uncertainty, assumptions, conservatism and judgement applied in their assessments. A threat to the efficacy of risk decision-making is presented by the inconsistent communication of these aspects (13a, 14a, 15a, 16a).

In the case of Challenger (case study 1), there was dispute over which evidence was admissible in the final decision to launch - judgement-based perspectives were dismissed because there was no supporting experimental evidence or “launch commit criteria” established for the historical damage observed on the O-ring seals (Vaughan, 1996, p308). In addition, the only full scale vehicle temperature-related data of the SRB performance was perceived to show no correlation between temperature and the O-ring erosion and blow-by. In the Lamington viaduct case (case study 2), risk analysis identified the viaduct as high risk but the lack of meaningful intervention or timely resource allocation suggests either a lack of a risk threshold, or a lack of accountability to those thresholds. Instead, mitigations in the form of annual inspections provided an alternative acceptability measure for decision-makers. In the military aviation maintenance case (case study 3), the conspicuous lack of evidence of formal risk analysis infers the use of judgement about the acceptability of the partial repair undertaken. Extant ‘risk to life’ thresholds were not evidenced directly, but were seemingly not exceeded relative to the decision-maker’s perspective because the decision to continue flying the aircraft was not referred to a nominated duty holder, as required by military aviation maintenance policy (Military Aviation Authority, 2015).

4.3.3.4 Acceptability of risk assessment output

Differences of opinion are also apparent regarding the acceptability of structural integrity risk assessment output. Survey respondents demonstrated trust in their

risk processes used (survey Question 17, 22, 26, 30, 34), but the credibility of these beliefs is undermined by widespread inconsistencies in the survey responses. For example, Question 30 identified a majority of respondents agreeing that their organisation's approach to structural integrity risk maintains a transparent audit trail. Inconsistent respondents stated this agreement despite indicating difficulty when establishing clear likelihoods (11), not communicating the use of uncertainties (13a), use of assumptions (14a), use of conservatism (15a), or subjective judgement (16a), and believing that decisions in operational situations are not clearly justified (22). Reviewing across Questions 11, 12, 13a, 14a, 15a, 16a and 22, this behaviour was shown by 20 (58.8%) participants. Some participants were repeatedly inconsistent. Table 4.3 shows the prevalence of the inconsistency by question, and Table 4.4 shows the breakdown of these inconsistencies by role. Over half of respondents in every role and category exhibited this behaviour. This could be the result of organisational influences (culture, training), the individual (hindsight bias, a lack of feedback resulting in presumed association bias, overconfidence), or due to the survey (e.g. ambiguous wording). The resolution offered by Question 30 is insufficient to identify the cause of this result and more research is necessary to explore whether the behaviour is explainable by a *more often than not*-type perspective on whether risk management methods allows transparency of risk analyses.

Further concerns regarding the efficacy of structural integrity risk assessment outputs stem from respondents who cited the use of independence, assurance activity, other people's consultation and review, and their methods for data handling as reasons to support their confidence (34). These support mechanisms are not necessarily available in operational situations under restricted time, information or resource. For example, access to advice may be limited by the decision horizon or communication links, and assurance activity may only target higher levels of abstractions, as suggested by Participant 61 (aviation, operator/end user, risk analyst) (35):

> *"...it's not formal assurance, other than the independent safety advisor, when people came in to so called assure us, they were looking at*

Question	Proportion of contradictions	n
Q30 | Q11	32.4%	11/34
Q30 | Q12	3%	2/33
Q30 | Q13a	14%	4/29
Q30 | Q14a	16%	5/31
Q30 | Q15a	15%	5/33
Q30 | Q16a	30%	10/33
Q30 | Q22	18%	6/34

Table 4.3: The prevalence of inconsistencies regarding transparency in risk assessment approaches by survey respondents, shown by question number. *n* is the number of responses to the dependent question that were inconsistent.

		n	Proportion of group
Category	Designer/equipment manufacturer	5	62.5%
	Operator/end user	8	57.1%
	Regulatory/policy maker	6	60.0%
Role	Risk decision-maker	10	58.8%
	Advisor	5	62.5%
	Risk analyst	4	50.0%
	Commentator	1	100.0%

Table 4.4: Proportion of respondents in each role and organisation category that exhibited inconsistencies between to Question 30 and earlier questions. *n* is the number of respondents from that group that demonstrated inconsistency.

> *process compliance rather than reviewing our actual risk analysis, the decision we came to..."*

Assurance mechanisms could also give erroneous confirmation that risk management practices were correctly carried-out, although, it would be inappropriate to speculate on the generality of assurance effectiveness given the possible variations across respondent organisation assurance mechanisms and perceptions on this activity.

Perceptions about risk assessment use by decision-makers suggest the

existence of cognitive bias toward decision-maker ability, and motivational bias about decision-maker handling. Participant recall about previous experiences of good decisions made by decision-makers (26), may only reflect good outcomes:

> Participant 31 (aviation, operator/end user, risk decision-maker) - *"Over many years of both witnessing and completing the analysis, I have not witnessed a decision that has been assessed and it's prescribed actions failed to manage the associated risks."*(26)

Positive outcomes may have resulted from poor decisions particularly in systems with redundancy and complex interactions, and where catastrophic failures are rare (Dekker, 2019, p268, p397). In addition, decision-maker familiarity with the subject domain may misdirect the decision choice, as Participant 15 (aviation, regulatory/policy maker, advisor) commented (26):

> *"By their very nature, SI risks (indeed, any IM[21] risks) are but one of hundreds of individual risks held by the ADH[22] - thus attempting to persuade ADHs to commit funding to mitigate a particular SI risk means that it must compete, on an equal footing, with 'operating risks' (pilot training, Aircrew Eqpt etc) that the ADH will be far more familiar with."*

Evidence of disagreement about the acceptability of the risk assessment output is also seen in case studies 1 and 2. Engineers who had made judgements about the risk to Challenger from the cold weather had difficulty in justifying their risk perception to other stakeholders (case study 1). In the viaduct case, the risk analyses that found Lamington as a high risk did not convince decision-makers of the need to take action (case study 2).

[21] Integrity Management.

[22] Aviation Duty Holder, a title given to a nominated individual responsible for risk-to-life decision-making, see Military Aviation Authority (2020a).

4.3.4 Theme 3 - Situational constraints

Case studies were selected on the basis of operational situations that constrain decision-maker analysis and intervention options. Survey responses were framed by a need to consider answers relative to "operational situations"[23]. This enabled specific consideration about how naturalistic environments influence the handling of risky decisions.

4.3.4.1 Situation compels handling approach

When a decision is constrained by ecological factors specific to a given situation, decision-makers are known to be adaptive to time pressures and complexity by either accelerating their processes, filtering only a subset of the information available, or by adjusting decision strategies entirely (Payne et al., 1993, p38). Practitioner divergence from normative analytical prescriptions under real-world conditions is known about (Mosier, 2008, p45). The survey results are consistent with these conditions. The use of judgement is widespread to support decisions or justify the acceptability of a safe threshold in their given situation (survey Question 9, 16, 21a) and extracts from participant free text responses highlight the reliance upon judgement as a consequence of naturalistic constraints[24]:

> Participant 12: *"In structures we do not have analysis methods to calculate risk, the analysis has been answered as judgemental methods, supported by static and fatigue calculations."* (18)
>
> Participant 31: *"...'Clear' consequence and likelihood are extremely subjective, applied judgement and experience will often determine the threshold (as opposed to defined, measurable metrics)..."* (18)
>
> Participant 51: *"There is a tendency to rely on past experiences - for example a crack occurs to a part of the airframe and the often the first thing asked is - has this been seen before and what did we do then.*

[23] Defined in the survey as "a situation that is time critical, as well as information poor and/or resource constrained." See Appendix D.

[24] Participant 12 - aviation, designer/equipment manufacturer, risk analyst; Participant 31 - aviation, operator/end user, risk decision-maker; Participant 51 - aviation, operator/end user, risk decision-maker.

We have to rely on the corporate memory of those around us..." (18)

The case studies are also consistent with these results. The unavailability of supporting resources, information or analysis created a distinguishable context from decisions made with the support of adequate analysis. In case 3, no analysis supported the perception that an incomplete repair would be safe for the subsequent flights, and in case 1 the MTI engineers plugged the gap in available empirical data (evidence of the whole shuttle system performance at low temperatures) with their judgements. In both situations the judgements proffered were not able to be supported by analysis within the decision horizon.

Specifically, practitioners had difficulty with generating likelihoods (11, case 1, case 3), establishing or assessing the relevance of a pre-determined safe threshold for structural faults (9e, case 2) and experienced influencing forces or pressures when deliberating structural risks (25, case 1, case 2). Specific adaptive behaviour is not identifiable from the survey data, but citing environmental factors or practical constraints as necessary information for decision-making (19) infers the bearing that situational constraints have upon practitioners during risk deliberation.

The broad agreement regarding the use of conservatism when handling structural integrity risk (15) is another context-driven practice. The uncertain precision that is a consequence of conservatism (Aven, 2016a) may be appropriate while the conditions appropriate to the use of conservatism remain valid. However, for enduring issues that have long lead times to resolution, such as fault repairs awaiting resources or formal analyses, conservatism can mask the true extent of risk through assumptions. The validity of the conservative choice only remains so while the assumptions remain relevant and appropriate. This may help short term actions, but cloud longer term planning decisions owing to a lack of precision about the true extent of a particular risky prospect's nature.

That conservatism is used is indicative of the nature of structural integrity risk assessment and subsequent risk management activites. An inability to be precise within the time frame necessary for a decision may result in unnecessary intervention, commit resource wastefully, or may lead to greater attention being

paid to risks that may be subsequently considered arbitrary. Risk theorists argue against the use of conservatism, but pragmatically speaking, the only option available to safety practitioners may be to justify uncertain and imprecise choices with a risk averse approach.

4.3.4.2 Suitability of methods for the situation

Risk analysis method choice appears to be predisposed upon information, advice, time or resources available in a situation. Some survey respondents argued for deterministic treatment of structural fault assessments only (survey Question 9, 20), some apply conservative factors in their treatment of structural issues (15), and some have difficulty in generating a clear likelihood (11). Differences between respondent roles and categories in their method choices, although not significant, suggests that the decision-making context is a predictor of method suitability. For example the dominance of 'what if' and 'safety audit' qualitative methods by operator/end users may be a consequence of usability or portability. The variability of methods applied to structural integrity risks across survey participants highlights a potential source of tension, particularly if risk analyses are completed in differing forms (qualitative/quantitative/a mixture of both). Risk analysis methods predetermined by published procedures may not be suitable for all circumstances in which practitioners are expected to apply them. Consequently, any procedural requirements for handling risk or defining safe thresholds should be cognizant of a decision-maker's situational constraints before advocating a particular approach.

4.3.4.3 Sources of bias

Situational influences may mislead decision-makers through cognitive bias effects. Influencing pressures (survey Question 25) can introduce motivational biases through social pressure (Montibeller and von Winterfeldt, 2015) and have been shown to distort decision-making strategies toward more heuristic strategies (Hoffman, 2009, p177). Similarly, time stress (21a) leads to more risk seeking behaviour (Wickens et al. (2013, p274), Naweed and Kingshott (2019)). Feedback

from previous decisions (26, 34) is also suggestive of a potential source for post hoc fallacy (Fulkerson, 1996, p103) or hindsight bias (Baybutt, 2018).

Other motivational biases are also promoted by situational availability of resources or information. Authority or courtesy biases (Baybutt, 2018, p206) may lead practitioners to attribute greater weight or accuracy to opinions from more senior colleagues or more experienced advisors, who may be the only peer support available (21c).

Social pressure may also explain the inconsistencies displayed by some participants (30). Peer pressure or bandwagon effects (Baybutt, 2018, p207) encourage participants within a social group to conform with the majority. In organisations with highly visible risk practices (4, 7) and assurance practices that review compliance (35), this may generate sufficient social influence on individuals to trigger such effects.

4.3.5 Theme 4 - The use of judgement

Throughout the previous three themes, the use of and dependence upon judgement has been inferred as a consequence of the situational constraints or owing to different perspectives on what is possible regarding structural integrity risk. Survey participants outlined their use of judgement in assessing information (survey Question 19) and in their process forming a decision (20) about a structural integrity risk. One participant noted that judgement is *expected* to be used in risk assessments about structural integrity:

> Participant 40 (aviation, regulator/policy maker, risk decision-maker): *"It is accepted this type of risk analysis is not an exact science so SQEP judgement, assumptions etc all play a part in developing a strong safety argument."*(18)

Judgement use in structural integrity risk analysis was not restricted by respondent role or category, although each group identified different modal responses for the circumstances in which they use their judgement (21a). For instance, operator/end users most cited reason for using judgement was "when

there is a lack of available expertise", whereas designer/equipment manufacturer's modal responses were jointly "when I believe that technical information is conservative" and "when there is insufficient information available to support an objective analysis".

The use of judgement in relation to structural risks affects all roles and organisation categories polled (16). The impact of judgement on risk decisions made relative to a threshold was discussed in Section 4.3.3.3, but this may be further affected by the difficulties practitioners noted in applying their judgement to structural integrity risk decisions (21c). Respondents believed they would be better supported with better access to *knowledge*, improvements in the accessibility of useful *information*, and improvements in *analysis* techniques. While these are desirable, they do not necessarily reflect situational limitations, such as access to information or a means of generating analysis within the time available. However, Participant 34 (aviation, operator/end user, risk decision-maker) noted (21d):

> *"The scenarios during which judgement is used tend to be operationally focused therefore there is very little that can be done other than ensure that adequate professional training and career experience have been accrued by those being asked to make those judgement recommendations."*

4.3.5.1 Relating engineering judgement to types of thinking

Without the luxury of analytical support, decision-makers deliberating risky prospects and the information available to them are making *unsupported* deliberative judgements, Type 2 processing (see Section 2.3.1). Taking the perspective of dual process theory (Section 2.3), the distinction of unsupported decisions can be more usefully categorised using recognised terms to reflect the situational constraints identified as effecting structural integrity risk decisions. Type 1 and 2 processing modes are not mutually exclusive in respect to risk thinking. Despite risk analysis aiming to deliver an objectively rational assessment of a hazard's future prospects and reducing human subjectivity, subjectively chosen

parameters or modelling criteria, for example, introduce scope for cognitive error in traditional risk methods (Aven, 2018a, p242-243). Holistic subtleties such as reversibility, persistence, ubiquity, tolerability, equity, catastrophic potential, controllability, or voluntariness (van Asselt and Renn, 2011, p442) and cognitive characteristics like "affect"[25] also have bearing on the conduct of risk assessment, decision-making, communication and management (Slovic et al., 2004, p314). Accounting for human effects in risk assessment is therefore appropriate for completeness.

Although distinguishing between Type 1 and Type 2 is not necessarily important for considering the "unconditional risk" (Aven, 2018a, p242) of a fault or failure in a decision situation, understanding the influence and improving the transparency of a decision base can benefit the epistemological footing of a judgement-based argument about a risky decision (Aven, 2018a, p242).

Figures 4.27 and 4.28 visualise the conceptual interaction between decision-maker processing, resource and time that describes the situations reviewed in the case studies (case study 1, 2 (on-call track decision), and 3). The development of the concept was presented in Green et al. (2021) (reproduced at Appendix F). Situations with the resources, information and time to enable the formation of analytical assessments are referred to as a "supported" Type 2 risk assessment. Under operational constraints, however, an "unsupported" condition may arise when decision-makers are compelled to form a decision but their judgement lacks the backing of an analytical assessment. A deliberative assessment may still be possible, reliant upon the decision-maker's ability to act on the available information. Figure 4.27 depicts Type 2 thinking for a risk assessment that has been able to make use of modelling, analysis, deterministic solutions, historical information, or another credible and rational tool appropriate to a given decision. Figure 4.28 shows the unsupported condition, which lacks the use of such reference materials (labels are consistent with Figure 4.27).

Actions taken in each case study show the compelling effect that judgement

[25] Affect means "the specific quality of "goodness" or "badness" (1) experienced as a feeling state (with or without consciousness) and (2) demarcating a positive or negative quality of a stimulus" (Slovic et al., 2004, p312).

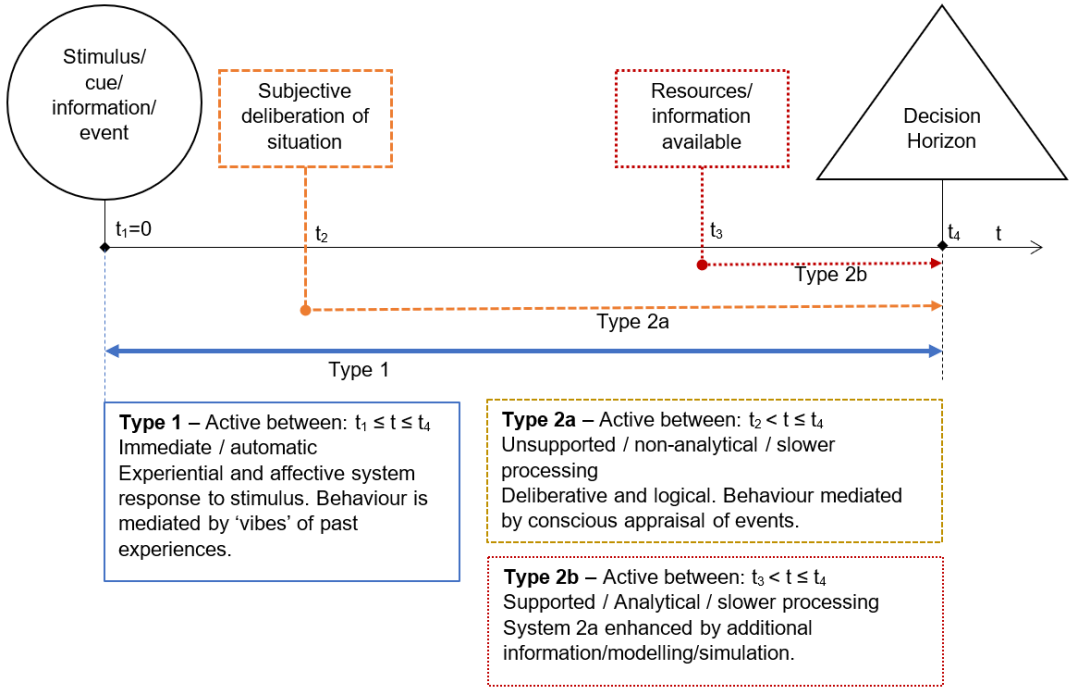


Figure 4.27: Decision follows a 'supported' Type 2 thought process.

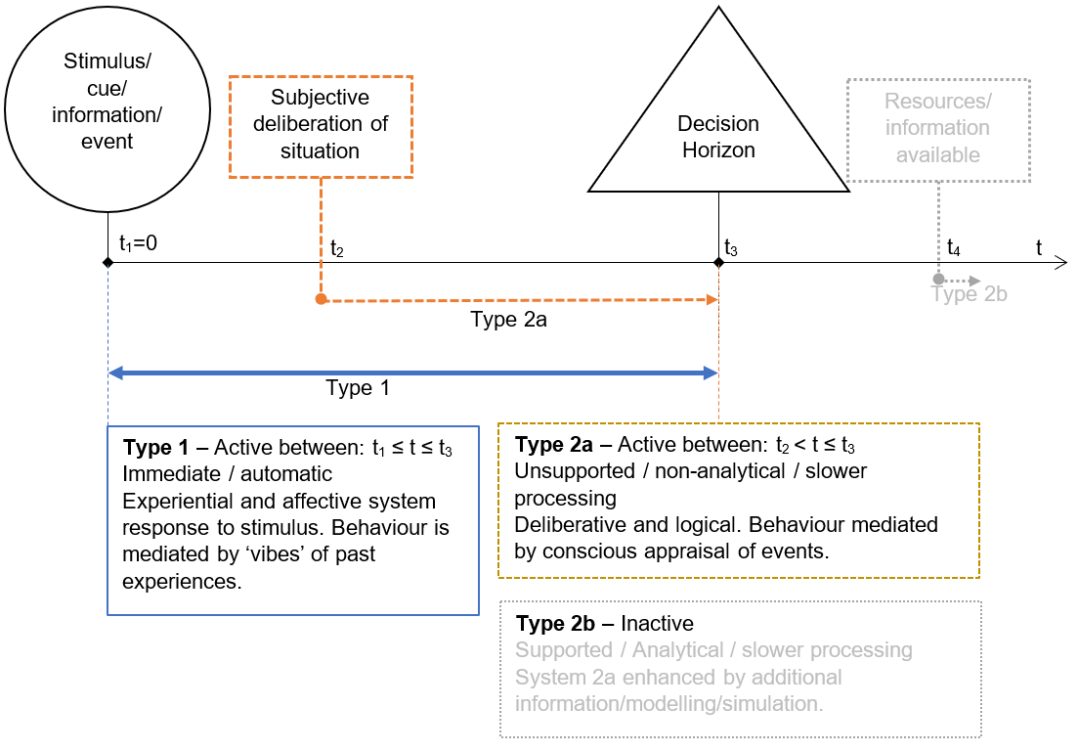


Figure 4.28: Decision follows an 'unsupported' Type 2 thought process.

has on behaviour as a consequence of goal conflicts within the situation and of the local rationality of the protagonists in each situation (Dekker, 2019, p324). All cases show examples where reliance upon unsupported judgement was necessary to decision-making. In Challenger's case (case study 1), the organisation was not configured to consider the MTI engineer's judgement-based deliberations, which led to the dismissal of their claims. Their holistic argument that the ambient temperature conflicted with the available database for the complete vehicle was weaker than the competing rationally-presented argument concerning the data. Had the advocates of 'no launch' been able to present their argument more coherently, they may have gained more support.

This need for better justification of judgement is also identifiable in case 3, where the engineering manager had not explained the grounds for deferral. No logical bridge is apparent between the available information (damage dimensions, the non-destructive test for subsurface damage indications and the use of a recognised repair technique) and the claim that deferral was appropriate. Although links can be inferred, the perception of acceptability depends on the knowledge-base of the inferencer. For instance, where the deferral wording reads "NDT carried out with no fault indications", there is an assumption that the NDT was 100% accurate, which may not be true and can be open to challenge.

The literature review coupled with the survey responses and case studies illustrate that judgement-use is not formally regulated or appropriately operationalised for practitioners in the domains studied. Improvements in the ability to make defensible, logical and transparent arguments in time-critical situations where hard evidence and further analysis are unavailable is vital to bolster the quality of such unsupported judgements and the decisions which they inform.

4.3.6 The need for improved support for structural integrity risk decision-makers

The results show that judgement use is widespread across structural integrity risk practitioners. The management of structural integrity risk is a valued activity,

however the treatment of structural faults and failures using risk assessment methods is contested amongst practitioners. This disagreement appears limited to whether it is appropriate to treat structural risks probabilistically or deterministically. In operational situations where an engineering structural integrity risk threatens the safe operation of in-service equipment, the freedoms of the decision-maker to access information and resource are sometimes constrained and the time to deliberate the issue can be limited. Individual perceptions can also encourage reliance upon judgement, for example where there is disagreement over the relevance of some information to a given situation.

Boundedly rational decision-makers will use their judgement as they deem most appropriate for a given structural integrity risk decision in a given situation. For some organisations, in particular within the aviation domain, this creates tension with promulgated procedures for handling risks, such as risk thresholds for the safe use of equipment, as discussed in themes 2, 3 and 4.

There has been no guidance identified within aviation regulations as to how engineering decision-makers should be expected to apply their judgement under such time-critical operational circumstances. While this may be considered appropriate policy practice (to only explain the ends, not necessarily the means) by affording flexibility for the regulated community in operational circumstances (case 3 and Military Aviation Authority (2020b, p8)), the survey results illustrate individualistic opinions towards structural integrity risk, the opportunity for bias and variable situational constraints that may lead to performance variation, and inconsistent application of safe practices when handling structural integrity risks. Extant risk management requirements may outline expectations for decision-making, but there lacks specific guidance on how best to apply individual judgement to a given situation.

Where decision-makers are compelled to make a choice and they do not have access to normative decision-support mechanisms, there is an opportunity for improvement. If personal unsupported judgements can be formally and systematically incorporated into risk management activity, and justifications can be documented more transparently and made more open to scrutiny, decision-makers

may benefit from more timely feedback via metacognition and peer review prior to a decision, and through post-hoc learning and expertise development. This is exemplified by the circumstances seen in case study 1. Decision-makers are deficient of an appropriate support mechanism to guide the use of judgement for handling structural integrity risk. The definition of such a mechanism may also reduce the tension experienced between meeting organisation risk requirements and doing what is achievable in a given situation.

4.4 Chapter summary

Primary and secondary data has been collected to fulfil research objectives, rigorously identify weaknesses in real-world structural integrity risk management practices, and inform the development of a framework for supporting decision-makers deliberating structural faults in operational situations.

Responses from a survey of practitioners with experience of structural integrity risk, were combined with analysis from three case studies that were representative of operational maintenance decision-making about structural faults. Results coalesced around four themes relevant to structural integrity risk, regarding: the usefulness of risk concepts in practical settings; differing opinions; situational constraints; and the use of judgement. In particular, although risk concepts are valued in practical settings, the approaches applied to structural integrity faults are heavily subjective. There was widespread use of the judgement by practitioners in operational situations, especially when decision-makers are compelled to make a choice about a structural fault.

Chapter 5

Novel framework methodology

Engineering knowledge can be classified into three components: scientific knowledge, tested under controlled conditions; the application of approximating assumptions to the design of engineering systems, proven by actual performance and supported by testing in idealised conditions; and 'rules of thumb' or heuristics (Blockley, 1992, p16). The application of engineering knowledge to problems boils down to understanding the usefulness or appropriateness of a particular method to a problem (Blockley, 1992, p17). In the time-constrained context defining this research, a determination of system failure risk given a structural fault will only be possible if the required information and correct model for calculating risk is available within the decision horizon. For aircraft maintenance contexts, unless testing has been carried-out and *a priori* assumptions have been made regarding a particular fault manifestation, this information will be unavailable. Clearly, it is unreasonable to expect this for every potential structural fault that could occur throughout the lifetime of an aircraft. This leaves knowledge established by 'rules of thumb' as the only option available to a decision actor considering a structural fault in the operational aircraft maintenance context. This mode of knowledge application suffers from individualistic bias and means that the problem must be looked at through a different lens with subjectivity at its heart.

This chapter describes the methodological approach to developing and testing a novel framework for supporting aviation maintenance decision-makers compelled

to make a decision under operational conditions regarding a structural fault.

5.1 Novel framework development (Phase 3)

The need for a new approach to support structural integrity risk decision-makers during aircraft maintenance under operational conditions was identifiable from the literature review and survey/case study results.

Results from these two work strands provided the requirement sets for a new approach, but emerged at different times during the research project as a consequence of unforeseen delays in gaining ethical approval (see Section 3.1.2 and Appendix A). As a result, existing risk analysis methods in the literature were scrutinised first using a holistic set of *contextual* requirements. The Phase 2 survey outcomes followed, initially establishing a set of *user* requirements, which were then developed into *system* requirements. Ideally, a single set of requirements would have been more efficient, but this practical shortfall was mitigated by checking the consistency of the contextual requirements following analysis of the Phase 2 survey results. The internal consistency can be seen by comparing the list of contextual requirements with the list of user requirements in Section 5.1.2.

5.1.1 Existing risk handling methods

Reviewing the applicability of risk approaches already in the literature was necessary to verify the need for a novel method to support aviation engineering managers deliberating structural integrity risks under operational constraints. The inconsistency between the literature (which pays little attention to judgement-based risk decision-making for aircraft maintenance settings) and the real-world practitioner actions identified in Chapter 4 infers an answer to this is a necessity. This inconsistency may be indicative of a lack of coherence between academia and practitioner domains regarding safety science approaches (Rae et al., 2020). *Contextual* requirements outlined at the start of Chapter 2 were based upon the

author's experience in the environment, and enabled scrutiny of candidate methods identified during the literature review.

Consequently, existing risk analysis and decision methods eliminated were those that were purely quantitative, those that necessitated model formulation, those that required simulation, and those that were dependent upon expert information input or structured elicitation. These were on account of requirements for usability by generalist aircraft engineering managers, the time available to make a decision, and information or resources available to derive solutions. Supportive methods were found to be those that were consistent with systems considerations, resilience engineering approaches, heuristic-based approaches and those that were cognisant of naturalistic decision-making environment. Although some methods seemed functionally suitable (usually as a result of their ontology) their value in operational conditions (specifically under time and resource constraints) meant that a more ecologically compatible tool was deemed necessary to support individual decision-makers apply their judgement to aircraft structural integrity maintenance risks. To achieve this necessitated incorporating the Phase 2 results.

5.1.2 Structural integrity practitioners survey results

Inductive analysis of survey responses generated a specific set of *system* requirements, which were used to develop and validate a new methodology for its suitability and verify the scrutiny applied to existing methods.

5.1.2.1 Verifying existing method use

Survey participants were asked to identify processes they use in support of SI risk analysis or decision-making, and to reflect on risk parameters (likelihood and consequence) and appraisal aspects (uncertainties, assumptions, conservatism, and use of judgement) relative to structural integrity risk analyses (survey Question 9d, Chapter 4). Responses to these questions (in particular by participants identifying as an "operator/end user") verified that the contextual requirements established for scrutinising and excluding existing risk analysis methods for

suitability against aircraft structural integrity risks under operational constraints were appropriate (Section 4.1.2).

5.1.2.2 User and system requirements

In order to design a new framework in a rigorous manner, a systems approach was followed using practitioner survey data as the input user (or stakeholder) requirements. Following a structured approach provided *a priori* assurance that the end framework would be acceptable for its purpose before testing in Phase 4. Each survey question enabled a statement to be made about how practitioners responded to an aspect of in-service structural integrity risk analysis; these were the user requirements (Appendix I, Table I.1). In raw form, the user requirements could be interpreted in a number of ways, so broader system requirements were established by identifying common themes and patterns from the survey data, which formed the *system* requirements[1] (Faisandier et al., 2020). After arranging all relevant Phase 2 survey data into the established theme, each theme was further thematically analysed to provide more detailed system characteristics against which the new methodology could be measured and revised in the event of deficiency (Figure 5.1). The higher-order themes alongside their respective sub-themes are laid out in Table 5.1. Maintaining traceability from the system requirements ensured that user requirements were matched to a particular feature in the design solution. The complete trace from each system requirement theme to the novel framework design features can be found in Appendix I.

[1] System requirements differ from user requirements because they establish the design and architecture of the system in order to fulfil the user requirements. For example, analysis of survey Question 8 (Section 4.1.2) found that participant organisations use both numerical and non-numerical representations of structural risk although the situation can obstruct the formation of quantitative assessments (Chapter 4). The system architecture themes emerging from this question relate to *external support, cues, justification* and *model form*.

System Requirements		*Development Model x-check*	
		v8	**Remarks**
Alternate Choices	Choices	Y	Choices are dictated by the heuristic choices at each gate. Although not objective, choice thresholds at each gate are individualistically driven and provide the framework to justify choice thresholds. Gate outcomes result in unambiguous alternatives being selected.
	Justify	Y	Alternate choices are defined by the decision archetype, specifying either fly or no fly, with a requirement to justify the choice if the situation requires. The argument form can be exploited to describe influences on judgement validity through rebuttals (i.e. factors that, if changed, would alter or reduce the strength of the argument). Assmt of speculative events and disregard of low P events is inherent to any argument, e.g. argue to "fly" because of warrant meaning low P event is disregarded.
	What is the hierarchy of choices?	Partly	Established by the fast and frugal tree in order: need to use judgement (because judgement is bias and less objective meaning errors are not easily attributable); basic test of judgement argument soundness; debias/consensus for judgement. Robustness is established regarding the argument at tertiary stage through debias/consensus but **physical system robustness (e.g. by recognising redundancy in a system or introducing performance limitations) are not explicitly ruled upon by the framework.** This is to enable flexibility .
	Dynamic aspects	Y	Document version control enables the assessment to be revised, **although this is not planned to be tested.**
External support	Justify need for support/escalation	Partly	This features within the alternatives presented, if the user answers "no" at Gate 2 or 3. Although, the **"wide reaching ramifications" perspective is a consideration for the holistic handling of the risk (other aircraft affected)**.
	Use of external assessment - justify use/trust	Y	If external assessment is available, the framework prompts users to consider the argument basis at Gates 1 and 2 (signified by the dotted lines to indicate the need to justify is discretionary and dependent upon the situation).
	Enable inclusion of external support (qualitative or quan)	Y	In addition to enabling users to consider external support as a "supported" assessment and therefore not requiring judgement (Gate 1 decision option), external support may also be considered within the formation of argument to provide warrants or backing.

Figure 5.1: Screen capture of system requirement cross-checking during novel framework development

Table 5.1: Themes and sub-themes within user requirements.

Theme	**Sub-theme**
Test Strategy	Target audience to prove new method.
	Show general behavioural tendencies.
Applicability	Project definition.
	Conditions of use: not strategic decisions or emergency decisions; framed by sociotechnical system.
Generality	Operational decision-making centric.
	Flexible to context.
	Handle prospective situations.
	Communicate non-objective factors.
	Generalised process: gather information/situational awareness/define the problem; interim action/generating and considering options; intervention/outcome review.
Holistic Process	Prompt for either holistic risk management or individual treatment.
	Focus on judgement on risk.
	Trigger for use of framework.
	Follow generalised decision process. Support the wider risk and decision-making domain (alignment with literature on process).
	Structure the model for simplicity and learning (simple, memorable, easy to revisit) - visual, use mnemonic or metaphor, alliteration, etc.
	Able to be applied by non-specialists in operational contexts.
	Enables auditability and transparency of the decision basis (independent understanding, non-SME understanding, revisitable for learning).
Output	Display/information presentation (time sensitive information: info available, info acted upon. Qualifiers and rebuttals: assumptions, uncertainties, conservatisms).
	Justification: escalation requirements; threshold for action; reasonability of action.
Alternate Choices	Choices.
	Justify alternate selection.

Table 5.1: (continued)

Theme	Sub-theme
	Hierarchy of choices.
	Dynamic aspects.
External Support	Justify need for support/escalation.
	Use of external assessment - justify use/trust.
	Enable inclusion of external support (qualitative or quantitative).
Cues	Capture individual reaction.
	Structure cues, label, and the conscious use of cues.
	Presenting/displaying cue use.
Justification	Mechanisms for justification.
	Individual judgement justification.
Feelings and Influences	Decision-maker values.
	Situational influence on intuition.
	Mechanisms to support.
Model Form	Provide links between: information, cues, choice; risk-decision (values, agreement); situation capture - actor perspective - actor honesty; decision - responsibility; actor frame - information available; information available - reflection on source; situational information - judgement; risk, judgement, true risk; model, supported RA, true risk.
	Inherent features (threshold for action; Handle risk modes (numeric/verbal); decision-centric; framing information appropriate to structural faults; include judgement and experience; justification; transparency; confidence; handle subjective and non-subjective).
	Discrete features (show information available and situation specifics; identify use of uncertainty, assumptions, conservatism and judgement; output is a decision(not a risk); simple structure and ease of recall; generalisability of situation clear; initial filter stage for escalation).

5.1.3 New framework concept

The ontological approaches for decision-making that appeal to the research context from the literature review are those of *naturalistic decision making* and *fast-and-frugal heuristics*, which complement the developing safety science concept of *resilience*. These approaches suit the operational decision-making domain by simplifying the decision problem as far as appropriate[2], essential if any framework is to be useful to a decision-maker under constraints of time, resource and information.

5.1.3.1 Naturalistic decision-making

Naturalistic decision-making (NDM) seeks to understand the nature of real-world decision-making, particularly how actors apply their experience to real-world settings. Works in this field argue that traditional normative decision-analysis approaches, such as quantified expressions of expected loss, are seldom applied in operational settings (Yang and Haugen, 2015). While a variety of perspectives on the characteristics of naturalistic decision-making scenarios exist (Gore et al., 2015, p223), a common characteristic is that the solution derived by a human for the decision problem in real-world scenarios is situationally dependent. The degrees of freedom available for decision-making are governed by the situation through dictating temporal constraints, controlling the resources available to the decision-maker, and by being cognitively limited by the decision-maker's own abilities. Survey participant responses indicated difficulty in providing a likelihood for structural integrity risks (survey Question 11), inferring that normative risk analysis methods are not always suitable, at least for making decisions upon structural integrity risk where the decision outputs may be inconsistent across decision-makers (see Section 4.3 for a full discussion).

[2]This is not to suggest that a reductionist approach is utilised. Reductionist mechanisms, such as fault and event trees, create problems for resilient systems by failing to account for the dynamic and adaptive characteristics, a basis of systems theory (Grant et al., 2018, p100). The aim of simplifying the decision problem here is to generalise the framework as far as possible to make it adaptable to operational situations, and to improve the usability of the framework for the practitioners it is intended to support.

NDM methods provide a suitable theoretical footing upon which to found a new supportive framework for decision-makers by being cognisant of the situational aspects. Conveniently, NDM also complements the developing safety paradigm of *resilience engineering*, which aims to help people deal with complex situations under pressure, without employing reductionist strategies (Dekker, 2019, p402-403) as exemplified by fault tree, Bayesian network or multi-criteria decision methods. The applicability of these perspectives is strengthened by the case study analysis, which demonstrated that some situations compel a decision to be made, *unsupported* by analysis. This places expectancy upon decision-makers to use their judgement and experience to combine the available information and arrive at a decision.

Attending to any structural integrity risk decision problem in a naturalistic manner acknowledges that decision-makers are compelled to make a decision one way or another, irrespective of their access to an analytical solution. The decision will be formed from the available information that they have access to in their situation and their own knowledge and experiences. This assumption necessitates a different view on decision rationality. In traditional decision analysis methods, normative methods comply with probability axioms and use rules for rational choice, maximising utility or identifying dominating alternatives. Pursuing a method that is sympathetic to situational constraints requires a boundedly rational perspective on the situation (Simon, 1955), which assumes rationality is limited by the environmental constraints and the decision-maker's cognition. This shift makes the decision problem inherently subjective and dependent upon decision-maker perception, and formally includes individualistic judgement as a conditional factor in the choice.

5.1.3.2 Judgement, heuristics and justification through argument

Shifting the focus of a decision support mechanism toward the perspective of the decision-maker moves towards subjective territory and away from the usually objective rule-based means of analytical deduction that aircraft maintenance (Naweed and Kingshott, 2019, p82) and engineering in general (Blockley, 1992,

p454) typifies. Other professional fields already operate in this subjective space, such as lawyers arguing a case on behalf of a client (Toulmin (1958, p15-16), Sinnot-Armstrong and Fogelin (2015, p353-382), Gainsburg et al. (2016, p335)), historians making claims about past events from limited data (Govier, 2010), or medical professionals (Osmo and Benbenishty (2004), Gainsburg et al. (2016, p337)). Gainsburg et al. (2016, p334) highlights that structural engineering design processes are reliant on subjective perspectives: "judgment is always involved in engineering design; the engineer's goal is to reduce the risks associated with subjectivity and uncertainty." Descriptive decision analysis has systematically identified that humans are fallible when they use subjective judgement. Normative decision approaches, the idealised intent for risk analysis, are designed to reduce the effects of bias and fallacy from human sources, by acting upon objective evidence and applying deductive, rational analysis to optimise the decision. Nevertheless, the survey results (Section 4.1.2, Question 16) indicate that decision-makers operating under an existing risk management regime use their unsupported judgement to make decisions on structural integrity risks in spite of normative methods. This is backed by claims in the literature of analyst values influencing risk assessments (Durodié (2017), Montibeller and von Winterfeldt (2015), Slovic et al. (2004)).

The survey results also indicated that a proportion of participants do not communicate when they have applied judgement in their risk assessment (Section 4.1.2, survey Question 16a), and to a lesser degree other risk appraisal aspects such as assumptions, uncertainties and conservatism (Section 4.1.2, 13a to 15a). Clear communication when applying a risk method is essential (Rae et al., 2012) for achieving appropriate governance of risk activities (van Asselt and Renn, 2011), establishing credibility (Durodié, 2017) and making clear the reasons behind a decision judgement (Hardman and Ayton, 1997). Risk matrices are prevalent in industrial practice but are not configured to communicate factors like assumptions or uncertainties inherent within risk assessment (Cox, 2008). Hence, in organisations where regulatory and policy control of risk handling activities pervade and opaque risk tools are applied, the need for transparency

is currently not complied with. This could be considered as probative blindness (Rae and Alexander, 2017, p163) and a hazard in itself. Considering risk analysis transparency from a temporal viewpoint, safety management practices would benefit from a transparent record of justification at the time of the decision. Such perspectives would enable reflection on the material circumstances of the decision to target improved a priori support to decision-makers. Without such a perspective, judging a decision once the outcome is known is prejudiced and has the substantial benefit of hindsight, which clouds the accuracy of recall and has potential to mislead conclusions surrounding the defensibility of the decision basis (Baybutt, 2018).

Considering the complex nature of aircraft structural integrity risks, the unwieldy deterministic representations discussed in Chapter 2, and the requirement for a tailored response means that whole system representations will be inaccurate (Brugnach et al., 2010), particularly given the time-frame available to a decision-maker under operational constraints. A new framework should incorporate mechanisms that complement the use of judgement, but reduce the influence of subjective errors of omission and commission in a decision problem characterised by contextual constraints.

To the author's knowledge, no tool currently exists to support aircraft engineering decision-makers in using their judgement to deliberate structural risks. Other authors have also noted a lack of research addressing how engineers actually make decisions (Vermillion et al. (2015), Gernand (2018)). To consistently guide decision-makers in their application of judgement requires an approach that is flexible to the situation without being dependent upon the generation of analytical solutions that will only be available beyond the decision horizon.

Fast-and-frugal heuristics Fast-and-frugal heuristic methods can be applied to constrained decision situations through *search*, *stop* and *decision* rules that rely upon evolved human capacities such as memory, recognition, counting and recall (Todd and Gigerenzer, 2012). However, the approach relies upon identifying appropriate heuristics for particular domains by employing the validity of the

information cues used for decision-making, which is not viable for structural integrity problems in aircraft maintenance. This is for several reasons. Firstly, validating whether a particular cue relevant to structural integrity risk is a credible predictor of equipment failure is not economically viable for whole aircraft systems. Such an undertaking for even a single fault would necessitate an uneconomical testing or simulation burden. Secondly, natural variation in structure creates scatter in structural fatigue performance, a reason why structural designs attract safety factors (Theofanous, 2003). Consequently, the same variation means that classifying cues relevant to a structural fault as “signal” (rather than “noise”) cannot necessarily be assumed to be consistent across structural situations if applying a general risk decision heuristic (in the same way as envisaged in Hafenbrädl et al. (2016), for example). Thirdly, the information that would be available to a decision-maker under operational constraints may vary between contexts. Lastly, subjective factors prevalent in dynamic operational situations, such as the importance of the next mission introduce subjective safety perspectives to the decision that have bearing on the reasonability of a decision (Ale et al., 2015). This moves away from the objective evidence that fast-and-frugal heuristics require in which to compare validity across cues.

Notwithstanding, the principles of heuristic methods are attractive because they are simple to apply, transparent, fast in application and cost-effective (Hafenbrädl et al., 2016). They are anticipated to be usable in operational situations by generalist practitioners deliberating structural integrity risk decisions. The use of *search*, *stop* and *decision* rules can still be adopted for subjective consideration of a situation. However, this will generate an output that is boundedly rational relative to the user, instead of providing an ecologically rational output as intended by fast-and-frugal heuristic methods (Todd and Gigerenzer, 2012). Standardising such an approach would be considered procedurally rational (vice substantively rational) (Simon, 1979) - the framework prescribes a rational procedure, but relies upon the decision-maker to select their choice.

Fast-and-frugal trees are an application of heuristics to decision trees with rapid termination in a sequence of binary choices. One example uses the “take-the-best”

heuristic in a process of: ordered *search* of cues and information; a one-reason *stopping* rule on which to terminate a search of the cues; and a one-reason *decision-making* rule (Gigerenzer, 2005). The use of "take-the-best" is appropriate in situations which are *non-compensatory*[3]. For situations with compensatory cues, an alternative heuristic called "tallying", which involves counting cues in favour and against a particular proposition, is suggested (Gigerenzer, 2005). Fast-and-frugal trees have been applied successfully to medicine for triaging patients, ignoring probabilities and instead asking yes-or-no questions to resolve the decision outcome (Marewski and Gigerenzer, 2012). However, in these cases, the cues were able to be validated objectively owing to timely feedback loops. Aircraft accidents are, fortunately, rare but do not have the luxury of feedback loops.

Figure 1.1 simplified the aircraft maintenance decision situation faced in Figure 1.2, such that the decision is a binary choice that can be applied to a fast-and-frugal tree: fly the aircraft (with damage in-situ); or ground the aircraft (and repair, replace, or seek advice). This presents a choice between a certain alternative (ground) and a risky alternative (fly). Traditionally, including a probability p of an accident resulting from the structural damage (Figure 1.1) allows the problem to be rationalised using normative methods, such as by predicting consequences and factoring these by p. The resulting risk factor generated can then be compared against acceptable risk thresholds to deduce the action necessary. But if a meaningful quantification of risk cannot be generated, as indicated by the results from the survey analysis (Section 4.1.2, survey Question 11), then representing the problem through an alternate abstraction away from quantified measures of uncertainty can provide criteria for search, stopping and decision-making rules. An argument construction offers this functionality.

Argument Attending to rules of logic through verbal reasoning, argument combines information of different types, outside of mathematical operators (Hardman and Ayton, 1997). This is attractive in situations typified by mixed

[3]Decision strategies that make trade-offs between cues and attributes are compensatory, strategies that do not make trade-offs are non-compensatory (Payne et al., 1993, p29).

information availability. For example, a probabilistic expression of a particular risk can be included as a premise within an argument, alongside premises reflecting interpretive states of belief in order to substantiate a claim regarding a decision choice (Hansson and Hirsch Hadorn, 2018). The probabilistic expression might not be trusted by a decision-maker based on some knowledge of the analysis approach, so could be discounted.

An important factor from the user and context requirements is the need to provide a means of justification and transparency to enable decision-maker accountability. Critical reasoning and argument provide well understood mechanisms for enabling justification (Gainsburg et al., 2016). Indeed, justification is a tool of argument (Sinnot-Armstrong and Fogelin, 2015) and its previous application to decision-making scenarios speaks of its practical utility (Ramiah and Banks (2015), Osmo and Benbenishty (2004), Polacsek et al. (2018)). By enabling transparent justification, arguments could be peer-reviewed, challenged and improved before commitment to an intervention. Providing operational decision-makers with a usable framework that encodes such adaptive decision practice is also supportive of resilience engineering ideas. Supporting decision-makers as a component within a complex aircraft system, in which all states of operation cannot be foreseen (and therefore controlled) (Dekker, 2019, p402), accounts for the conditional work-arounds in which decisions must be made in some situations, in spite of organisational policies. In risk management, work-arounds may appear valid to decision actors, but hiding assumptions within a parochial and subjective risk assessment, or forming a risk perception on an incomplete mental model[4] may be misleading. Structured application of argument has been applied to aviation safety cases[5] (Dekker (2019, p203), Military Aviation Authority (2019), Conlin et al. (2004)), but has not been found applied to assessing aircraft maintenance risks outside of informal discourse justifying

[4] A mental model is an individual's structure of a situation that reflects understanding of the system and how it will respond, and may be created spontaneously or formed through training (Wickens et al., 2013, p236).

[5] A safety case is "a structured argument, supported by a body of evidence that provides a compelling, comprehensible and valid case that a system is safe for a given application in a given operating environment" (Military Aviation Authority, 2019, p7).

choices. Structuring decision justifications by rules of argument was anticipated to provide alternative means to assess the decision situation away from traditional risk characterisation. It has the potential to enable transparency, with the added benefit of providing a time-stamped record to benefit learning from reflection or post-hoc analysis, for example during a formal incident inquiry. The situation decomposition involved in forming an argument also complements the results from survey Question 21b, where a majority of structural integrity practitioners decompose their judgements about risk (see Section 4.1.3).

5.1.3.3 Forming the heuristic decision tree

The operation of the framework can be viewed as an analogy where the heuristic decision tree provides the road map, and argument providing the vehicle. The decision tree fixes the alternatives possible to “fly” or “ground”, focussing the decision-maker toward the tasks of deliberatively assessing available information, preferences, logic of the situation and immediate choice. Argument then provides the means with which to arrange the information being acted upon to deliver the judgement-based decision.

The design of the decision tree was formed by reflecting on the results of the practitioner survey, particularly Question 21a[6]. Reducing the decision-making burden by increasing decision process simplicity was anticipated to provide focus on justifying the safety decision at hand. The decision tree is arranged into three Gates, through which the user passes if they answer “yes” at each stage, visually presented in Figure 5.2.

[6]Question 21 required participants to answer “Do you ever use judgement when making or supporting a decision regarding in-service structural risks?”, and for participants that answered *Yes* was followed by Question 21a: “In which types of situation do you use judgement?”, which provided a selection of answers and an “Other (specify)” option.

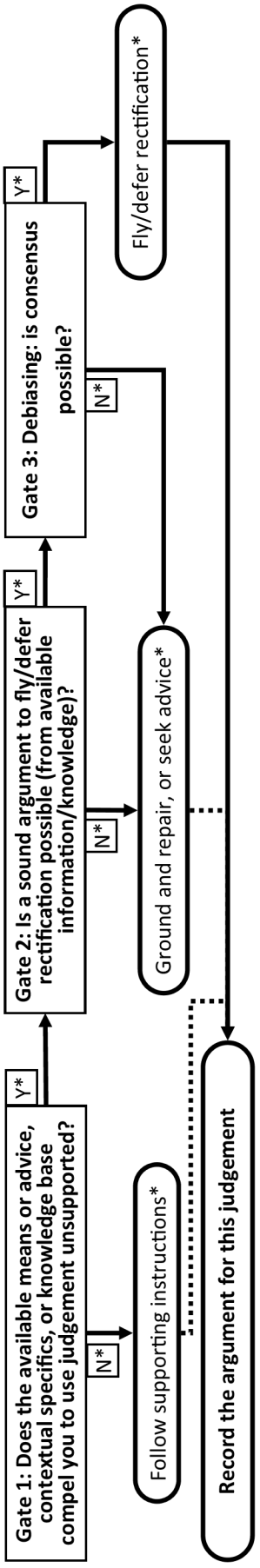


Figure 5.2: Novel framework heuristic decision tree, indicating the three "Gates".

Gate 1 Knowing that judgement is prone to error in both experts and naive participants of descriptive decision studies (Bazerman and Moore (2009), Richie and Josephson (2018), Hardman and Ayton (1997), Croskerry et al. (2013a)), the dominating consideration is whether judgement must be used in the situation. Yates et al. (2011) found in naive situations that despite legal requirements, individualistic policies regarding seatbelt use were dictated by situational factors such as time of the journey, journey length and whether they were carrying a passenger. In the absence of any domain-specific evidence, the first Gate attempts to referee such behaviour by requiring consideration of whether judgement use is necessary to be used by the decision-maker in the situation. The wording of Gate 1 was developed by identifying the broad themes from survey Question 21a, to capture the circumstances in which practitioners are compelled to use their judgement in as few words as necessary. The intent behind this was to maintain accessibility of the framework, without resorting to a checklist approach.

> **Gate 1:** Does the available means or advice, contextual specifics, or knowledge base compel you to use judgement unsupported?

If the user answers “No” to Gate 1, then there exists some reason (either objectively or subjectively perceived) why the decision actor does not need to rely on their judgement, given the context or the information available to them. The decision termination from this predictor instructs that those defaulting reasons are followed to provide the solution to the structural fault being considered at that moment in time, rather than intuitively resorting to judgement.

Gate 2 With the survey having established that quantifiable likelihood measures of risks are invariably not possible (Section 4.1.2, survey Question 11), Gate 2 must be flexible to circumstances that require a threshold for acceptability based on, in the worst cases, a qualitative risk representation. This is a necessity to ensure consistency with risk practices advocated by aviation regulation (Military Aviation Authority (2018a), Civil Aviation Authority (2015)). Gate 2 provides a flexible means of incorporating policy-derived quantitative and qualitative thresholds, by evaluating

how *sound* the decision-maker's argument is to continue flying an aircraft that has a particular structural fault. Argument has previously been advocated as a tool for either complementing or substituting for probabilistic risk analysis (Hansson and Hirsch Hadorn, 2018). It has also been used to substantiate decision-making in health domains (Krause et al. (1995), Hardman and Ayton (1997)), social services (Osmo and Benbenishty, 2004), industrial labour dispute decisions (Ramiah and Banks, 2015), engineering design choices (Polacsek et al., 2018) and discussed as being used informally throughout structural engineering design processes (Gainsburg et al., 2016, p333-335).

The output from Gate 2 is an inductive argument[7] about the risk of a structural fault, made by the decision-maker that makes use of information available to them within the temporal and resource constraints of the decision situation. This approach is suited to naturalistic scenarios, within which it has been found humans tend to unconsciously adopt adaptive solutions that are incompatible with normative rationality (Maldonato et al. (2017), Payne et al. (1993, p38)). Decision-makers are thus provided with a means to substantiate their boundedly rational perspective, lending evidential importance to intrinsic capacities like, experience and expertise (Hurteau et al., 2020), memory (Marewski et al., 2010), and recognition (Cohen et al., 1996) to help justify complex risk decision problems. Other authors have advocated a means of including such perceptive information sources into risk analysis (Aven (2018b), Askeland et al. (2017), Mousavi and Gigerenzer (2014), Pidgeon (1998), van Asselt and Renn (2011), Osmo and Benbenishty (2004)). Structuring a risk assessment in this manner negates the need to force unrealistic assumptions in generating a precise probabilistic measure for risk, an approach previously applied in Hardman and Ayton (1997).

The ultimate aim of an argument is a "sound" argument (Groarke, 2020), but since this classification formally applies to deductive arguments, an analogous

[7]There is a philosophically complicated distinction between inductive and deductive logic, on both of which arguments can be founded. A full discussion is beyond the scope of the thesis. For brevity, an inductive argument is one where the truth of the premises does not guarantee the conclusion (Fulkerson, 1996, p11), as might be considered when viewing a risky problem with a chance of success (not a gurantee). A deductive argument, where the truth of the premises guarantees the conclusion, which is drawn from formal logic (Fulkerson, 1996, p10). This definition is adequate for the purposes of outlining the mechanics of this thesis.

measure of argument acceptability is necessary for the ordinary language and reasoning found in natural discourse. Arguments in natural forms use informal logic and the evaluative measures for such are the acceptability (A), relevance (R) and sufficiency (S) (Groarke (2020), Govier (2010, p87)[8]) of the premises and claim, or the ARS criteria. Any shortfall in the argument identified through the ARS criteria requires addressing by the decision-maker and if unable to be appropriately addressed, the aircraft should not fly on account of a flaw in the argument. This is a precautionary approach, which is appropriate for safety settings in order to minimise the chances of harm in the face of uncertainty (Conlin et al. (2004), Bodde et al. (2018)). As in Gate 1, the decision rule for Gate 2 is also consistent with a "one-reason" decision rule attending to the "take-the-best" heuristic. Other authors advocate rules or taxonomies for classifying argument assertions that oppose or support an overall claim, in a tallying-style form (Krause et al. (1995), Hardman and Ayton (1997)). However, a tallying approach to combining argument propositions was considered not appropriate because the cues available from a newly emergent fault situation under operational constraints are unlikely to be validated as compensatory (Gigerenzer, 2005). Using a one-reason rule reduces ambiguity and provides the clear rule-based threshold for Gate 2:

> **Gate 2:** Is a sound argument to fly/defer rectification possible (from available information/knowledge)?

If a sound argument cannot be made, the decision alternative selected for the aircraft is "ground and repair, or seek advice". Seek advice provides the option to identify additional information that may clarify uncertainties, validate assumptions or discount premises that prevent a sound argument to fly from being formed. If a sound argument can be formed, then the decision-maker can move on to Gate 3. Discussion regarding documenting of the argument about the risky prospect is below.

Gate 3 The final gate acts upon the findings from descriptive decision analysis that humans are prone to bias, suffering from innate or hardwired biases, and

[8]Govier (2010) refers to the "S" (sufficiency) criteria as "G" for good grounds.

acquiring biases, which are associated with heuristic forms of decision-making (Croskerry et al., 2013a). Furthermore, it is well known that domain experts are not immune to bias (Rae and Alexander, 2017, p160), and specifically, engineers have shown occurrences of ambiguity aversion (Brown and Utley, 2019). Although Gate 2 attempts to couple human intuitive judgement-based responses (Type 1 processing), with deliberative processing of the situation (Type 2 processing) in order to reduce the latent effects of bias, a decision-maker may be unable to prevent such bias even when it is known about (Croskerry et al., 2013b).

Consequently, Gate 3 formally introduces a second opinion to review and reflect on the argument basis. Such a perspective is expected to be readily available from peers or other practitioners within reach of the decision-maker in operational conditions. It is not expected that access to a large peer group would be possible, as is required by normative group elicitation methods (see Barons et al. (2018) for example) or for similar consideration of judgement-based approaches as in Aven (2018b).

A second opinion can enrich the information used in the decision-maker's argument to reduce biases, such as: by offering alternative grounds about risk influencing factors not initially considered (Yang and Haugen, 2016, p106); encouraging the consideration of non-confirmatory evidence that may have been overlooked owing to confirmatory bias (Baybutt, 2018); provide the decision-maker with an opportunity to partake in "perspective-taking", which has been shown to reduce the influence of wishful-thinking bias (Rose and Aspiras, 2020); and to help the decision-maker reflect and validate intuitions (Hurteau et al., 2020). Montibeller and von Winterfeldt (2015) reviewed biases found in risk analysis, and offered debiasing techniques. Those biases applicable to risk assessment[9] that may be reduced by interaction with a second opinion include:

– *Anchoring.* "Relying too much on the first piece of information" (Baybutt, 2018, p207). Use different experts who use different anchors (Montibeller and von Winterfeldt, 2015, p1233)

[9]Biases with descriptions relevant to parameter quantification are not discussed, although it is acknowledged there may be an analogous qualitative equivalent.

- *Availability/ease of recall/recency bias.* “giving more weight to recent info” (Day, 2020, Slide 11), “the probability of an event that is easily recalled is overstated” (Montibeller and von Winterfeldt, 2015, p1233), “people make judgements about the probability of events by the ease with which examples come to mind...how easily they can think of something similar” (Baybutt, 2018, p207). Provide counter examples (Montibeller and von Winterfeldt, 2015, p1233).

- *Myopic problem representation/either-or bias.* “Two choices only, cutting off creation of alternatives” (Day, 2020, Slide 11), “when an oversimplified problem representation is adopted based on an incomplete mental model of a decision problem” (Montibeller and von Winterfeldt, 2015, p1233). Explicitly encourage thinking about more objectives, new alternatives or other possible states of the future (Montibeller and von Winterfeldt, 2015, p1233).

- *Omission of important variables/information bias.* “Seeking information when it does not affect action” (Day, 2020, Slide 11), “when an important variable is overlooked” (Montibeller and von Winterfeldt, 2015, p1233). Prompt for alternatives and objectives, ask for extreme or unusual scenarios, and use group elicitation techniques (Montibeller and von Winterfeldt, 2015, p1233).

- *Overconfidence bias.* “Subjective confidence in one’s judgement is greater than the objective accuracy of the judgement” (Baybutt, 2018, p207), “feeling strongly about your opinion regardless of accuracy, less likely to seek out information” (Day, 2020, Slide 11), “when decision-makers provide estimates for a given parameter that are above the actual performance (overestimation) or when the range of variation they provide is too narrow (overprecision)” (Montibeller and von Winterfeldt, 2015, p1233). Use counterfactuals to challenge extremes (Montibeller and von Winterfeldt, 2015, p1233).

- *Affect influence/salience effect.* “Getting ‘hooked’ emotionally, over-focussing on vivid imagery and ignoring other information” (Day, 2020, Slide 11),

"emotion such as fear, pleasure etc. influences decisions" (Baybutt, 2018, p207), "when there is an emotional pre-dispoition for, or against, a specific outcome or option that taints judgements" (Montibeller and von Winterfeldt, 2015, p1235). Use multiple experts with alternative points of view.

– *Confirmation bias.* "Listening to and trusting only information that confirms your beliefs" (Day, 2020, Slide 11), "tendency to search for, interpret, favor, and recall information in a way that confirms one's pre-existing beliefs or prior knowledge, or aligns one's values while discounting or rejecting information that does not. Information may be remembered selectively or interpreted in a biased way" (Baybutt, 2018, p207), "When there is a desire to confirm one's belief, leading to unconscious selectivity in the acquisition and use of evidence" (Montibeller and von Winterfeldt, 2015, p1235). Use multiple experts with different points of view about hypotheses, use counterfactuals, probe for evidence for alternative hypotheses (Montibeller and von Winterfeldt, 2015, p1235).

– *Desirability of a positive event or consequence/survivorship bias.* "Looking for positive outcomes and neglecting what was lost" (Day, 2020, Slide 11), "when the desirability of an outcome leads to an increase in the extent to which it is expected to occur" (Montibeller and von Winterfeldt, 2015, p1235). Use multiple experts with alternative points of view (Montibeller and von Winterfeldt, 2015, p1235).

– *Undesirability of a negative event or consequence/ostrich bias.* "Avoiding information perceived as potentially unpleasant or harmful" (Day, 2020, Slide 11), "when there is a desire to be cautious, prudent or conservative in estimates that may be related to harmful consequences" (Montibeller and von Winterfeldt, 2015, p1235). Use multiple experts point of view (Montibeller and von Winterfeldt, 2015, p1235).

– *Desirability of options/choice supportive bias.* "Promoting an option you selected and/or demoting other options" (Day, 2020, Slide 11), "over- or

underestimating probabilities, consequences, values, or weights in a direction that favours a desired alternative" (Montibeller and von Winterfeldt, 2015, p1235). Use analysis with multiple stakeholders providing different value perspectives, use multiple experts with different opinions (Montibeller and von Winterfeldt, 2015, p1235).

Croskerry et al. (2013b) listed two strategies or environments suitable for debiasing: "workplace" and "educational". Gate 3 targets the former by helping address bias at the time of decision-making, providing a decision-support system with thresholds for acceptability (the one-reason decision rules) and controlling exposure (through Gate 1). As a consequence, Gate 3 seeks consensus with the argument described by the decision-maker in order to verify that a second opinion considers the decision argument to be sound.

Gate 3: Debiasing: is consensus possible?

If a second opinion provides consensus, it is reasonable to consider that the argument is sound given the resources available in the situation. Therefore, if consensus can be achieved, the decision to fly or defer rectification may be taken. If consensus cannot be achieved and new information brought by the second opinion cannot be accounted for within the argument to fly (under the same ARS criteria in Gate 2), the argument cannot be considered sound. Consequently, the decision alternative reverts to "ground and repair, or seek advice".

A summary of the heuristic search, stop and decide criteria for each gate is in Table 5.2.

Documenting an argument in-place of probabilistic representation of risk Gate 2 introduces the specific requirement for the decision-maker to consider the risky decision as an argument, by proclaiming a relationship between some evidence and a conclusion. An argument is a discourse that provides support for a conclusion (Fulkerson, 1996, p1-9), the effectiveness of which is related to its validity. Rather than expecting decision-makers to intrinsically determine the nature of an argument supporting or opposing flight given the structural fault, a

Gate	Heuristic criteria
1	**Search** for reasons, resulting from the means available, advice provided, contextual specifics or from within ones own knowledge that compel one to rely upon their own judgement without deferment to pre-established rules for action. **Stop** at the first cue that compels the use of ones judgement e.g. owing to a lack of information about a key variable. **Decide** if the use of one's judgement has been compelled, continue to Gate 2, else terminate and revert to pre-established rules for action.
2	**Search** for extrinsic and intrinsic cues that can be arranged into an argument regarding whether the aircraft is safe to fly or not. **Stop** once a sound argument, consistent with ARS criteria, is formed, and the level of abstraction is deemed appropriate for the context. **Decide** if a sound argument has been formed, continue to Gate 3, else terminate and elect to ground and repair the aircraft or seek advice (which may aid the generation of the sound argument).
3	**Search** for a suitable practitioner to act as a second opinion (within the limits of the situation), who themselves searches the situation for relevant cues and scrutinises the established argument at Gate 2 to provide challenge or support to the argument. **Stop** when all cues brought to bear by the second-opinion have been accounted for in the argument. **Decide** if consensus has been established then decide to fly or defer maintenance activity, else terminate and elect to ground and repair the aircraft, or seek advice.

Table 5.2: Expressions used to codify the case studies in identifying unsupported decision judgements to risky situations.

structured argument form has been woven into the framework to act as a guide, a prompt and a record. Structuring a risk decision justification by demanding deliberation of the information available has been called for (Brown, 2015), and an argument form offers several benefits. First, the deliberative allocation of information to the argument elements can prompt the decision-maker and aid in the self-scrutiny of their decision basis at the time: "understanding of justification relationships is needed as an input to metarecognition" (Cohen et al., 1996, p212). This can reduce the impact of optimism (Rose and Aspiras, 2020), engage Type 2 processing to reflect on intuitions (Hurteau et al., 2020, p333) and infer personal

accountability, which is a recognised strategy to help debias individuals (Croskerry et al., 2013b). Second, clear documentation of the reasoning enables transparent audit of the information relied upon for a decision (Krause et al., 1995), perhaps fulfilling regulatory requirements (Military Aviation Authority, 2020c) but additionally providing decision actors the opportunity to describe their "adaptive" perspective on the situation (Payne et al. (1993), Cohen et al. (1996)), and establishing an "apparent" validity of judgements made in the risk estimation (Rae and Alexander, 2017, p163). Third it provides a post hoc record for feedback to a decision-maker[10] (Wickens et al., 2013, p247-249) to enhance learning for novice development (Hurteau et al., 2020), enrich individual experiences by learning from mistakes (Phillips et al., 2005), and affirm individual memories of events in order to aid in subsequent heuristic use (Marewski et al., 2010). Fourth, it can mitigate the problem of distraction by forcing a systematic consideration of the information being acted upon (Payne et al., 1993, p181), a realistic problem in a dynamic operational environment. These benefits are attributable to both decision alternatives - fly or ground - since a documented argument can defend the credibility of the decision to fly, or create a justifiable defence of grounding an aircraft in the face of conflicting opinion and commentary.

With regards to risk assessments, using an argument form in place of a quantified analysis allows descriptive combination of cues (Hardman and Ayton, 1997), mitigating the need to consider redundant cues in isolation. For example where two cues may diagnostically indicate a particular symptom, such as a component's structural categorisation of importance and individual knowledge of the component function, both cues can be treated as contributory grounds towards a particular conclusion. In this form, the argument presents a fuller and more compelling justification of a risky decision in comparison to quantified measures alone, albeit outside of the form of traditional risk analyses. Suppressed premises, assumptions, uncertainties, conservatism or framing aspects of the

[10]Not all systems have strong feedback loops. Rarely occurring risks that propagate to failure are not likely to create the conditions were real failure feedback is possible. However, systems that are able to be effectively modelled may produce a simulated feedback opportunity or have access to experts with a strong understanding of physical laws in-lieu of real failure data, which may never emerge (Rae and Alexander, 2017, p159).

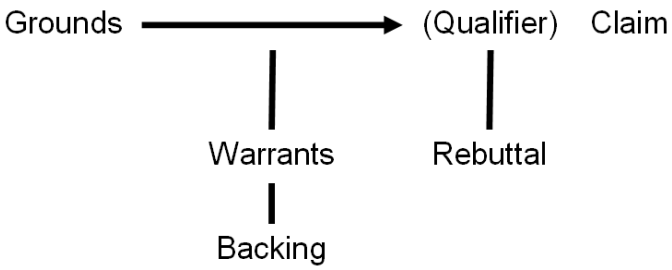


Figure 5.3: The Toulmin schema (Toulmin, 1958).

risky situation are less likely to be hidden from view in an argument form, because they form discrete elements in the logical steps from information being acted upon to a claim being made. The adaptive nature of the decision aid (Mosier, 2008, p51-53) facilitates the development of expertise, improves risk communication and aids metacognition. Indeed, better metacognitive function is believed to benefit information search, and encourage more analytical cognition and coherent judgement strategies (Mosier, 2008, p52).

The Toulmin schema (Toulmin, 1958) (Figure 5.3) was selected as the argument model owing to its roots in natural discourse (Groarke, 2020), its discrete functional elements that link practical reasoning to decisions to act (Ramiah and Banks, 2015), and its existing application to related study areas of aircraft engineering (Polacsek et al., 2018) and safety cases (Conlin et al., 2004). The schema describes a structuring of linguistic elements that make up an argument, such that the information within the argument is arranged to give logical appeal to a claim. In basic form, an argument consists of a proposition supported by at least one other proposition (Fulkerson, 1996, p4), but the Toulmin schema specifies 6 elements: the grounds; the claim; the warrants; the backing; the rebuttal; and the qualifier. These are defined (Toulmin (1958), Fulkerson (1996, p18-28)) and overlayed to an engineering maintenance context:

- *Grounds.* The visible, evidential or actionable information that is being used to act upon. The information that is argued *from*.

- *Claim.* The resulting claim about the safety of an aircraft and the decision selected. The assertion that is argued *to*.

- *Warrants.* The supporting evidence that provides an inference licence or authority to pass from the grounds to the claim.
- *Backing.* The information supporting the warrants that allows the warrants to be acted *in accordance with*.
- *Rebuttal.* Conditions of exception that would invalidate the *claim*, counter-arguments or counter-examples. Indications of the circumstances in which warrants may be set-aside. In a permissive sense, these would be considered assumptions.
- *Qualifier.* A subjective indication of the force or strength of the warrants enabling the claim.

To sustain transparency of meaning across decision actors, a local ordinal scale was established for the *qualifier* since probabilistic terms such as “improbable” already have associated quantitative meanings within some organisations (see for example Figure 2.1). A simple three-term scale was selected to be memorable but indicative in its purpose. Unlike Toulmin who embedded the qualifier within a linguistic argument phrasing, the form implied here provides a general classification of argument force/strength:

- *“Strong”.* A link between argument clauses is known to exist and is applicable to the situation being considered. There is a relevant causal chain or correlation that is accepted by consensus or authority.
- *“Some”.* The link is not strong, i.e. a link between the clauses is possible, but correlation or causality is not accepted by consensus or authority, or may be irrelevant to this situation.
- *“No”.* There is no clear link between the warrants able to be formed from available information, knowledge or data.

Gate 2 hence requires decision-makers to arrange the information they are acting upon into the relevant Toulmin elements of argument, although following

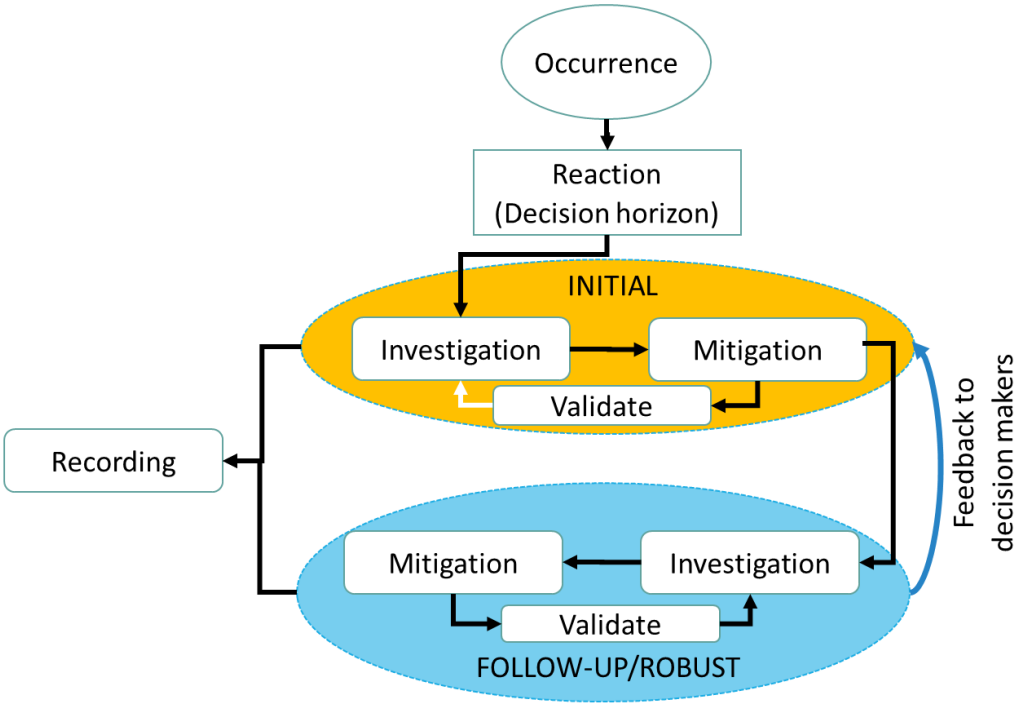


Figure 5.4: The initial judgement-based and follow-up decision-making processes in the treatment and adaptive management of a risk in an aviation maintenance setting.

pilot testing of the framework during Phase 4, the classical lexicon used by Toulmin was not easily related to by practitioners. The lexicon chosen became *Grounds* (Grounds), *Claim* (Claim), *Bridge* (Warrant), *Support* (Backing), *Voiding Conditions* (Rebuttal), and *Strength of Support* (Qualifier) in order to enhance usability by the intended audience. A visualisation of the framework is illustrated within Appendix L.1, Annex A.

Following an initial decision upon a structural fault, the documented baseline argument can be developed as more information is received, accounting for individual judgements and adaptive perspectives on the decision situation where objective information is poor, being updated as information and resources allow for improved knowledge on the phenomena under consideration. This is a temporal strategy that uses semi-closure of the problem to “extend a decision into the future by taking a provisory decision for implementation on a certain option to be reconsidered later on”, an adaptive management strategy (Hadorn, 2016, p224-225). This is represented in Figure 5.4.

5.1.3.4 Coherence

Alongside specific theoretical foundations, coherence with related standards demonstrate the acknowledgement of and effort to converge risk decision-making towards current accepted best practices. Divergence would only serve to further fragment a field which has struggled to maintain credibility (Section 2.2.2.1). With regards to decision-making, this is done through identifying how the framework incorporates the principles of decision quality. With regards to handling risk, this is done through identifying the relevance to ISO 31000.

Decision Quality Six elements describe the foundations of a good decision (Howard and Abbas, 2016): creative alternatives, useful information, sound reasoning, commitment to follow-through, a helpful frame, and clear values. All need to be present in order to distinguish between the quality of the decision and the quality of the outcome. In a judgement-based decision situation, bias and fallacy are inherent and may lead a decision-maker to believe they are making a quality decision, yet all the while suffering a bias, such as over-confidence toward their choice. The form of the framework provides an inherent structure in which to support elements of decision quality.

- *Useful information.* By construction of an argument, the usefulness of the information available to the decision is presentationally clear. Information may also be added at Gate 3 by the second-opinion.
- *Sound reasoning.* A structured argument layout is by its nature a platform for communicating the sound reasoning behind a decision.
- *Clear values.* Gate 1 enables decision-makers to distinguish between a situation that justifies the use of their judgement to make a decision, and one that does not. The structured argument layout enables these value perspectives of the decision-maker to be accounted for specific to the situation. For example, a typical value judgement in an aviation maintenance situation would be balancing operational urgency for the aircraft against the risk of structural fault leading to loss of the aircraft on that urgent flight.

- *Commitment to follow through.* The threshold criteria provided by the decision gates enables decision-makers to make an unambiguous choice. With argument providing the decomposition of the justification, the choice made is defensible and any counter-arguments can be critically scrutinised alongside the decision-maker's argument.

- *Creative alternatives.* The simplification of the decision problem (Figure 1.1) reduces the need for decision-makers to create alternatives for the immediate decision. Different courses of action may help resolve a decision to ground an aircraft, such as by acquiring resources or additional information, but the decision alternative fundamentally remains "is the aircraft allowed to be released for flight, or not?".

- *Helpful frame.* Four features of the framework help the decision-maker view the problem in a focussed manner. First, the decision required has been simplified (Figure 1.1) in order to reduce the restrict the opportunity for the decision-maker to create alternate frames, or ways of viewing the decision (Howard and Abbas, 2016, p37). Secondly, Gate 1 guides the framing of the decision by focussing on the decision-maker's situational perspective: is the aircraft required to fly the next mission? Who stands to benefit from the aircraft being permitted to fly? Is there a pressing time limit on the decision about the fault? Is missing technical information really necessary to make a justifiable decision on this fault? Thirdly, justification by argument enables self-reflection on the frame of the decision by decomposing the decision into contributing factors. Lastly, Gate 3 enables an alternative frame from a second-opinion in order to scrutinise the perspective of the decision-maker.

Risk standards - ISO 31000:2018 The set of guidelines established for risk management in The British Standards Institution (2018a) is generic in its approach, considering the scope of activities undertaken by an organisation and proactively addressing threats to objectives before they occur (The British Standards Institution, 2018a). The novel framework is consistent with the requirements of ISO

31000:2018, despite being designed from a reactive perspective. Risk identification activities are not expected to be undertaken since in the aircraft maintenance context, the hazard will generally already be manifest. The remaining elements of risk management are still required to be undertaken: analysis, evaluation, treatment, communicating, monitoring, recording and reporting. These elements of the risk management process advocated by ISO 31000:2018 are accommodated within the novel framework, by providing the means for decision-makers to assess risks when compelled to use their judgement, in a documented fashion that provides a clear selection of a decision alternative. Furthermore, the definition of likelihood accommodates qualitative and subjective approaches (The British Standards Institution, 2018a, p2).

5.2 Methodology for novel framework testing (Phase 4)

The testing strategy for Phase 4 aimed to identify: the naturalistic usability of the novel framework for making judgement-based decisions on aircraft structural integrity maintenance risks; and the audit effectiveness from a third party perspective. Testing aimed to identify if a structured argument approach provides benefit for decision-making and risk management over an unstructured control method, addressing Objective 2. In order to provide representative information about the suitability of the novel framework in a naturalistic decision-making setting, the testing was focussed on engineering maintenance decision situations concerning a structural fault where the decision-maker was compelled to make a choice. Experiments that include study on intuition should be conducted representatively (Hoffrage and Marewski, 2015, p155), although this test was not only examining intuitive responses by decision-makers. It was therefore necessary to recruit a specific cohort of experienced engineering managers as participants, who would understand descriptive accounts of aircraft maintenance structural faults, be able to provide a justifiable decision on the action they would take in

response, and critically review such decisions. Furthermore, it was not anticipated that typical engineering maintenance managers would understand the concepts used in the novel framework, so it was necessary to generate and deliver a short training package.

Phase 4 was split into two independent sub-phases. Sub-phase 4.1 subjected two cohorts of participants to simulated realistic but fictional aircraft maintenance decision scenarios: the control cohort using an unstructured discretionary means to detail and justify the decision made; the intervention cohort using the novel framework as a structured means to capture the decision justification. Sub-phase 4.2 required a third cohort of participants to compare control and intervention decision justifications from Sub-phase 4.1 to identify which was easier to understand, which was more transparent, and which gave a better decision justification.

5.2.1 Environmental constraints

Prior to the commencement of Phase 4, the global pandemic that resulted from the spread of CoVID 19 lead to a significant shift in expectation for how Phase 4 data collection could proceed. The initial intent was to conduct in-person focus groups, in order to deliver training and test the novel framework collectively and efficiently. However, contingency plans were enacted following changes to public legislation to mitigate the virus spread (Coronavirus Act 2020, 2020), which necessitated a wholesale translation of the training and data collection to a virtual delivery style. The uncertainty regarding duration and impact, and variability of travel and meeting restrictions across the UK meant that shifting the contingent virtual delivery to the primary delivery means was the most robust option with the highest chance of success. Although there was an increased risk in the ability to deliver effective in-person training, this could be mitigated as far as possible through the generation of a written training guide and video call training sessions with participants.

5.2.2 Data collection design

5.2.2.1 Phase 4 purpose

To support the theoretical improvements outlined in Phase 3, framework usability and prospective benefits were required to be validated by the anticipated end users. Consequently, the purposes of Phase 4 were:

- to test the usability of the novel framework with decision-makers deliberating an aircraft structural fault in naturalistic settings;
- to test whether the novel framework generates improvements in the auditability of judgement-based decisions about aircraft structural faults;
- and to conduct these tests under representative simulations of operational aircraft maintenance situations, thereby permitting the use of judgement by decision-makers in such circumstances.

5.2.2.2 Constraints

Aircraft structural faults constitute a high consequence risk situation. Testing was necessarily constrained to simulated situations as real-life testing of structural faults would be neither ethically sound or economically viable. Related to these reasons, the accuracy of the decisions recorded in Phase 4 was not assessed, since the case studies relate to representative faults where there is no stability to the situation, where neither a frequentist nor classical statistic for risk to the aircraft can be applied (Askeland et al., 2017) or indeed would be helpful (Rae and Alexander, 2017) for explaining an immediate risk decision. This perspective attends to the pragmatic ontology adopted for this research, whereby the true risk of a situation may never be known, which conflicts with a realist perspective, as Rae and Alexander (2017, p157) outlines:

> *"The realist view maintains that risk is a real, objective and quantifiable truth. The likelihood of an event in the future becomes the frequency of that event with the benefit of hindsight. Whilst we cannot know for*

certain how accurate risk estimates are at the time they are made, their accuracy may (at least in principle) be knowable at some point in the future."

An additional constraint was the need to recruit experienced, non-naive participants for Phase 4 testing. The naturalistic setting of the case study and focussed design of the novel framework necessitated recruiting participants who would understand the content of the cases and context of the engineering maintenance decision required. However, while the author had privileged access to industry professionals, the availability of participants was constrained by their willingness to participate. Subsequently, the number of participants recruited was not anticipated to be high and so experimental design proceeded with a target of 20 participants. This expectation was moderated by experiences from Phase 2 data collection.

5.2.2.3 Data requirements

Sub-phase 4.1 subjected participants to fictional but realistic textual simulations of engineering decision-making scenarios, with one cohort documenting their decision using an unstructured control text input, and another cohort documenting their decision using the structured novel framework text input. Sub-phase 4.2 subsequently required participants to each compare several paired decision justifications from Sub-phase 4.1 and specify, on Likert scales, their preferences between the approaches.

Experimental design The most appropriate method with which to trial the novel framework in a representative decision-making scenario, within the constraints, was by simulation. Textual simulations in the form of six fictional *test cases* outlining realistic aircraft engineering maintenance situations were developed and allowed control of the information available to participant decision-makers. The format also allowed for ease of distribution to participants via an online survey tool (Qualtrics, 2020), which served as the data collection tool. An internet-based

format had the added benefit of convenience for the participant. This choice was the best compromise to simulate a realistic scenario given the constraints on the research. In real-world situations, practitioners receive information in a staggered fashion across a variety of mediums, rather than in a single prose format at the same time as presented to them in the test. Real situations would also have other management issues competing for the decision-maker's cognitive resource, distractions and other stressors, as well as access to advice from professional colleagues.

As a consequence of being unable to realistically recreate the physical environment within which naturalistic decisions are made, a *soft* time limit was placed on participants by making a countdown timer visible during consideration of each test case. The time limit was considered soft because answers given beyond the 20-minute limit were still admitted, but participants were not informed of this. This was done to ensure that as many answers would be received from the limited sample size, and adherence to the time limit could be measured during results analysis.

An additional consequence of the experimental design constraints was that Gate 3 of the novel framework was not able to be tested. Given the difficulty experienced in recruiting a non-naive cohort, convenience was considered a priority for participants, so being able to arrange an additional party to role-play the Gate 3 second opinion was considered infeasible. In order to simulate the actions of Gate 3, participants were hence informed within scenario context information. See Appendix K:

> *"For the purposes of the case studies assume that, if necessary, you have a similarly experienced colleague (Mrs A. Bloggs) as yourself who is in agreement about your assessments."*

Independent variables The independent variables of the experiment are the decision-making participants and their experience levels, the scenario simulations, the decision approach employed (control or intervention) and the raters or assessors involved in Sub-phase 4.2.

Participant experience was further identifiable by: how long the participant had been a practicing engineering manager; the last time the participant used their judgement to support an engineering decision about structural integrity risk under operational constraints; and whether the participant had previously received training in how to make judgement-based decisions in the course of their employment. Decision-makers in Sub-phase 4.1 were put into pairs based on similar levels of experience. One member of each pair was randomly assigned to using the novel framework (intervention cohort) while the other was assigned to the control cohort.

Dependent variables Sub-phases 4.1 and 4.2 provide two distinct dependent outputs of the experiment: decision-maker assessments; and independent review of those decision-maker assessments. These outputs are used to test the hypothesis that a structured decision support framework can aid both decision-making and auditability of the decision basis. Sub-phase 4.2 identifies how well the risk decision can be audited by a competent third-party[11]: information must provide transparency of how the decision outcome was realised, and be understandable to the competent individual (Lathrop and Ezell, 2017, p193).

- Phase 4.1 variables: individual assessment of the situation
    - Decision-maker's perception of usability of the decision approach followed, measured by a likert scale for confidence in each test case, and holistically reflecting on all test cases by a Likert scale for usability in operational engineering environments.
    - Decision-maker's ability to self-scrutinize, measured by asking for comments on whether there were any aspects of their judgement that they had difficulty in communicating.
    - Decision alternative selected.

[11]The British Standards Institution (2018b, p26) provides discussion on the generic skills necessary for an auditor of management systems.

- Quality and quantity of given information considered necessary to use within the decision justification.
- Time taken to complete each test case.

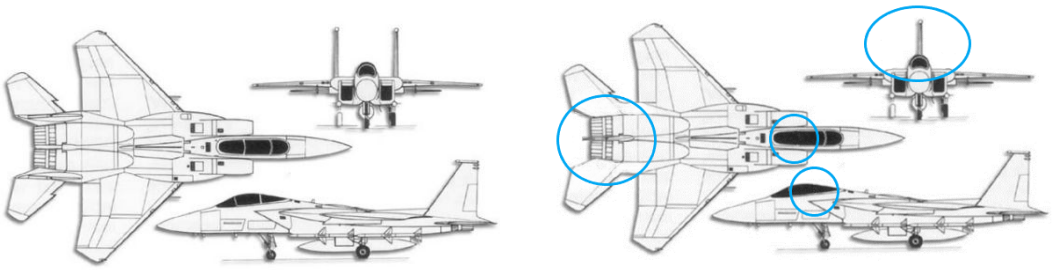
– Phase 4.2 variables: independent review for completeness of the risk decision assessment, through comparison of the control and intervention responses by asking:

- Which decision justification is easier to understand, and by how much.
- Which decision justification provides greater transparency of the decision basis, and by how much.
- Which decision justification offers a better justification, and by how much.

Test case design A fictional but realistic single-seat fighter aircraft was selected as the object of study. A fighter aircraft presents a very variable operational space, being subjected to a wide variety of environmental exposures, manouevre and equipment loads, and exposure to hazards in comparison to a transport aircraft. This was anticipated to encourage participants to recall knowledge and specific personal experiences to create a rich and variable data set for effective comparison of the decision justifications.

In order to increase realism and encourage comparable treatment to real-world decision scenarios, an existing fighter aircraft (McDonnell Douglas F-15 Eagle) was used as the base from which to create a fictional fighter aircraft. This was achieved by visually amending the general arrangement to create the appearance of an aircraft that had a single fin (by removing the two fin arrangement and replacing the image with a single central fin) and single crew occupancy (by removing canopy bracing structure to give the appearance of a single-piece canopy, as opposed to two discrete sections), see Figure 5.5. A fictional fleet context was provided to participants ahead of the individual test cases (Appendix K).

Test case content was devised from existing faults encountered during the author's employment and from true case studies identified earlier in the research,



(a) F15 general arrangement.

(b) Amended F15 general arrangement.

Figure 5.5: Comparison of the original F-15 general arrangement (5.5a), and the amended fighter presented to participants in Phase 4 testing (5.5b). Original image licensed under CC BY-SA (Zarco-Macross (Wikidot), 2013) and edited from the original by the author. Amendments indicated by circles.

but adjusted to remove any association with real events. Imagery was also created either as original content or amended from original imagery[12] to blend with the fictional aircraft. This was done to sustain realism with imagery that would typically be available to decision-makers, to decouple the image from its original content, protect its source, and minimise the opportunity where participants may be familiar with such damage mechanisms from their own experiences. This created a set of meta-real scenarios.

The test case information provided purposefully made the problem ambiguous and dependent upon decision-maker judgement within a short decision horizon. For example, by creating a conflict between advice or authoritative information, increasing the operational demand for the aircraft and the consequences of grounding the aircraft, and by introducing a delay to when authoritative advice might be received. Given the propensity for engineers to suffer ambiguity bias (Brown and Utley, 2019), this was expected to force participants to make clear their selection. Similarly, prompts were provided to stimulate individualistic consideration, for example test case 4 stated: "...and the dents are seemingly out of limits in accordance with maintenance manuals, as no limits are provided for this frame" (see Appendix K). The use of "seemingly" was deliberately ambiguous to encourage subjective deliberation.

[12]Permission was sought from the original content owners.

The experimental design controlled the decision recording mechanism, the simulated scenario, and the environmental effects, as far as possible, while aiming to retain realism of the scenario. This resulted in the decision justification being the only output from the participants. Every decision scenario necessitated at least one of two discrete judgements to be made: a judgement upon the risk of the fault to loss of the aircraft identified with the fault; an assessment of whether the fault found is analogous to other aircraft in the fleet. Test cases were presented to all participants in the order 1 to 6. This was expected to enable any within-experiment effects to be consistently recognised, for example familiarity with the test requirements, or for intervention participants fluency with the intervention method.

The information provided in the scenarios was consistent with the thematic findings from Question 19 of the survey (Section 4.1.3), and ensured that all cases contained contextual information regarding the operational environment, the demand/operational necessity to continue flying, the perceived time available, fault information, relevant fault and maintenance history, specific maintenance considerations, and how the fault related to published limits. For the remaining "essential" information fields (as determined by Phase 2 survey respondents), cases varied in their provisioning, as Table 5.3 indicates.

The cases were written to provide realism in a consistent format. Test cases 1 and 2 were checked by an experienced practitioner for realism and typicality, and acceptability was generalised to the remaining test cases[13].

5.2.2.4 Participant requirements

Phase 2 participant recruitment had revealed the difficulty in attracting practicing professional engineers to the research. As a consequence, Phase 4 experimental design was managed on the expectation of recruiting 20 participants, exploiting the same recruitment strategy as for Phase 2. Following dispatch of recruitment

[13]The availability of a practitioner to review the test cases restricted the ability to complete a full review of the test cases, but the limited review was sufficient to provide confidence in the typicality of the cases.

Information case class		Case no. 1	2	3	4	5	6	Coverage
Environmental factors	Influence of op. environment on practical/method decision aspects	3	3	2	4	2	2	100%
	Holistic operational necessity/need	2	3	3	2	4	5	100%
	Perceived time available	1	1	1	1	1	1	100%
Practical solution constraints	Published limits for component at risk	6	3	3	2	3	4	100%
	Maintenance considerations	7	5	2	10	4	5	100%
	Mitigation Options	6	9	6	7	5	10	100%
Measurable information, causal factors identification and vulnerability	Fault and relevant maintenance history	5	8	5	4	4	6	100%
	Damage/fault information	6	10	8	7	6	10	100%
	Usage data	1	6	6	0	2	1	83%
	Related fleet usage data	3	2	1	0	1	0	67%
	Privileged design data	4	1	0	0	0	1	50%
Individualistic factors	Awareness of information state	2	2	3	2	2	1	100%
	Individual judgement threshold	0	0	0	0	0	1	17%
	Need for advice	2	2	2	2	1	2	100%
	Human error consideration	1	0	0	0	1	0	33%
Risk aspects	Identification	3	2	7	6	3	3	100%
	Consequence focussed	0	3	2	2	0	1	67%
	Risk factor	0	0	0	0	0	0	0%

Table 5.3: Content analysis of test cases showing counts of information classes per case number, using main survey (Phase 2) Question 19 as a template, to indicate the information available to participants.

information, a deadline was set in order to proceed with systematic randomised sorting of participants into sub-phases and then cohorts. Time constraints had been limited by the requirement to seek further ethical approval from MODREC following initial approval for Phase 2 (Section 3.1.2). As a consequence, there was little flexibility within the research schedule to accommodate an extensive recruitment campaign, and participant allocation to their respective sub-phases continued beyond the deadline. Volunteers received after the deadline were allocated to Sub-phase 4.2, allowing Sub-phase 4.1 to progress unhindered from the deadline date.

Selection criteria In order to avoid capturing unhelpful data, strict qualifying and exclusion criteria were specified in order to ensure that only participants who would understand the engineering decision-making scenarios would be able to meaningfully partake. In addition to the Phase 2 criteria (Section 3.2.1) participants were required to have decision-making experience of aircraft structural integrity maintenance risks under operational constraints. Although restricting recruitment reduces the generalizability of the results, this was necessary given the ecological focus of the research and the need for understanding of the scenarios. Targeted recruitment also has benefits to reducing the influence of cultural or demographic differences which have been acknowledged to generate varying risk perceptions (Rae and Alexander, 2017, p158). No participants had to be discounted from the experiment.

- *Inclusion criteria.* Participants were required to be trained engineering managers with experience in the application of decision-making and the management of structural integrity risks, and have experience of decision-making on aircraft structural integrity engineering maintenance risks under operational constraints. This was to ensure they understood the context and nature of the test cases presented to them during testing.

- *Exclusion criteria.* Participants that were untrained engineering managers, had no exposure to structural risks or had no experience of aircraft structural

integrity risk decision-making under operational constraints were not invited to participate. Any false claims of suitability were expected to be easily identifiable through the quality and detail in the responses given in Sub-phase 4.1. Sub-phase 4.2 participants were only provided with selection boxes to complete their participation, so there was a risk that false claims of suitability may have escaped scrutiny. However, the targeted recruitment campaign, character of professionals, and contextual information provided ahead of the main data collection (Appendix K) were expected to sufficiently mitigate the chances of including unsuitable participants.

Recruitment method Survey respondents who had provided permission for subsequent contact were invited first[14], before opening the catchment wider through focal points within the UK MoD with a request for onward distribution in a hub and spoke-type approach (see Appendix J, Section J.2).

Incentivisation Monetary incentivisation was offered to volunteers as recompense for their time. In the worst case, the Sub-phase 4.1 intervention cohort would require pre-participation training in the novel framework before completing the decision scenarios, and would be expected to commit 4 hours; all other participants would be expected to commit approximately 2 hours. It was felt that a small monetary incentive would remunerate individuals expected to commit their time and cognitive effort.

Participant allocation Once recruited, participants were allocated a participant ID in the form “SIRM-Phase4-2020_n”, where *n* was a unique number.

Three sorts were required to allocate participants to their final role within the data collection. First, participants were randomly allocated to either Sub-phase 4.1 or Sub-phase 4.2, 10 participants to each. Next, Sub-phase 4.1 participants were paired up by similar experience levels and then within each pair randomly allocated to the control or intervention cohort. The scenarios allowed 20 minutes

[14]The email invitation to Phase 2 respondents included limited details of the results from Phase 2 to communicate the value their participation added and to incentivise further participation.

to produce a decision justification, which would take Sub-phase 4.1 participants a maximum of 2 hours to complete the 6 test cases. Finally, to ensure that all decision justifications from Sub-phase 4.1 received balanced scrutiny, an allocation plan for Sub-phase 4.2 was established. This distributed participants for Sub-phase 4.2 so that they would view at least one decision justification from each Sub-phase 4.1 pairing. Each Sub-phase 4.2 participant would only see one pairing for each case, to prevent any prior knowledge influencing the comparative assessment and maintain independence. It was estimated that each pairing review would take 20 minutes, so Sub-phase 4.2 participants were expected to commit 2 hours to complete their participation.

Initial sort After a preliminary sort to randomly allocate participants to Sub-phase 4.1 or 4.2, late volunteers necessitated a further sort, which concluded with $n = 10$ participants allocated to Sub-phase 4.1, and $n = 9$ to Sub-phase 4.2. Subsequent late recruitees beyond this initial sort were then allocated automatically to Sub-phase 4.2. Owing to the sequencing of Sub-phase 4.1 followed by Sub-phase 4.2, an operational decision to maximise Sub-phase 4.1 was appropriate to avoiding missing the opportunity to meet planned participant numbers for Sub-phase 4.1. Once allocated, participants were informed of when their participation was expected to commence and the deadline for completion. Participants were not informed of the existence of the other cohorts in order to avoid preconceptions in participants that may introduce bias and reduce realism.

Random sorting was carried-out using Microsoft Excel and the “RAND”, “RANK” and “INDEX” functions. A pseudo code for the random allocation is given in Algorithm 1, with the screenshot from Microsoft Excel in Figure 5.6.

Sub-phase 4.1 allocation It was expected that participants with more experience of making aircraft structural integrity maintenance decisions would provide more complete responses, as they would have more expertise in their field. Accordingly, participants allocated to Sub-phase 4.1 were requested by email to provide their experience:

Algorithm 1 Randomise allocation from "Participant ID list" to Sub-phases 4.1 and 4.2.

for all Rows containing Sub-phase position to be allocated **do**
  Assign random number (This step uses the "RAND" function)
end for
if row contains random number **then**
  Rank order random number from the array of random numbers created (This step uses the "RANK" function)
  if random number is nth largest **then**
    return Participant ID listed in nth row of $(1 \times n)$ array "participant ID list" (This step uses the "INDEX" function)
  end if
end if

Initial sort:

Participant ID list	Sub phase	Random number	Allocation (will keep changing)	Selection (13 Oct 20, 1146)
1	4.1	0.4953945	9	6
2	4.1	0.9381254	3	8
3	4.1	0.1631123	13	14
4	4.1	0.8098027	6	17
5	4.1	0.1117123	15	13
6	4.1	0.0996688	16	18
7	4.1	0.4507815	10	5
8	4.1	0.3753817	12	15
9	4.1		#N/A	#N/A
10	4.1		#N/A	#N/A
11	4.2	0.0387856	18	12
12	4.2	0.9667185	1	10
13	4.2	0.9419515	2	1
14	4.2	0.6564182	8	9
15	4.2	0.8352472	4	4
16	4.2	0.8210883	5	16
17	4.2	0.7197009	7	2
18	4.2	0.0544739	17	3
	4.2	0.4107218	11	11
	4.2	0.1416785	14	7

Follow-up sort (participant #19):

Participant ID list	Sub phase	Random number	Allocation (will keep changing)	Selection* (19 Oct 20, 1531)
1	4.1		#N/A	6
2	4.1		#N/A	8
3	4.1		#N/A	14
4	4.1		#N/A	17
7	4.1		#N/A	13
9	4.1		#N/A	18
10	4.1		#N/A	5
11	4.1		#N/A	15
12	4.1	0.970257896	1	10
16	4.1	0.003346748	19	16
19	4.2	0.578445078	4	19
	4.2	0.226651083	12	3
	4.2	0.690829688	3	1
	4.2	0.364496558	10	9
	4.2	0.780511995	2	7
	4.2	0.020150569	16	11
	4.2	0.390315417	9	12
	4.2	0.561379775	7	4
	4.2	0.240788867	11	2
	4.2		#N/A	19

* 8 slots of sub-phase 4.1 were already allocated on 13 Oct 20.

Figure 5.6: Microsoft Excel screenshot of Initial Sort of Phase 4 participants.

> "How many years of experience do you have in making operationally constrained decisions regarding aircraft engineering and maintenance, particularly regarding structural integrity risks/faults/issues?"

In order to prevent comparisons being dominated by the more experienced respondents, participants were ordered by experience before being assigned in to 5 pairs. Within each pair, each participant was then randomly allocated to the control or intervention cohort using Algorithm 1. Figure 5.7 illustrates the organisation of the pairing allocations to control and intervention cohorts, and their designation by capitalized letters V to Z. These pairings will be referred to as *experience pairs*.

Phase 4.1 Pairings	Participant ID (Pre-Sort)	Random No.	Control/Intervention	Participant ID (Post-Sort)	Sample (19 Oct 20, 1000) and 1615 (for V)	Phase 4.1 Pairings
V	10	0.44213296	Control	10	10	V
	16	0.21005896	Intervention	16	16	
W	15	0.87691075	Control	15	15	W
	17	0.34474872	Intervention	17	17	
X	8	0.97570252	Control	8	13	X
	13	0.88953321	Intervention	13	8	
Y	5	0.72452808	Control	5	14	Y
	14	0.69015425	Intervention	14	5	
Z	6	0.40900762	Control	18	18	Z
	18	0.42226887	Intervention	6	6	

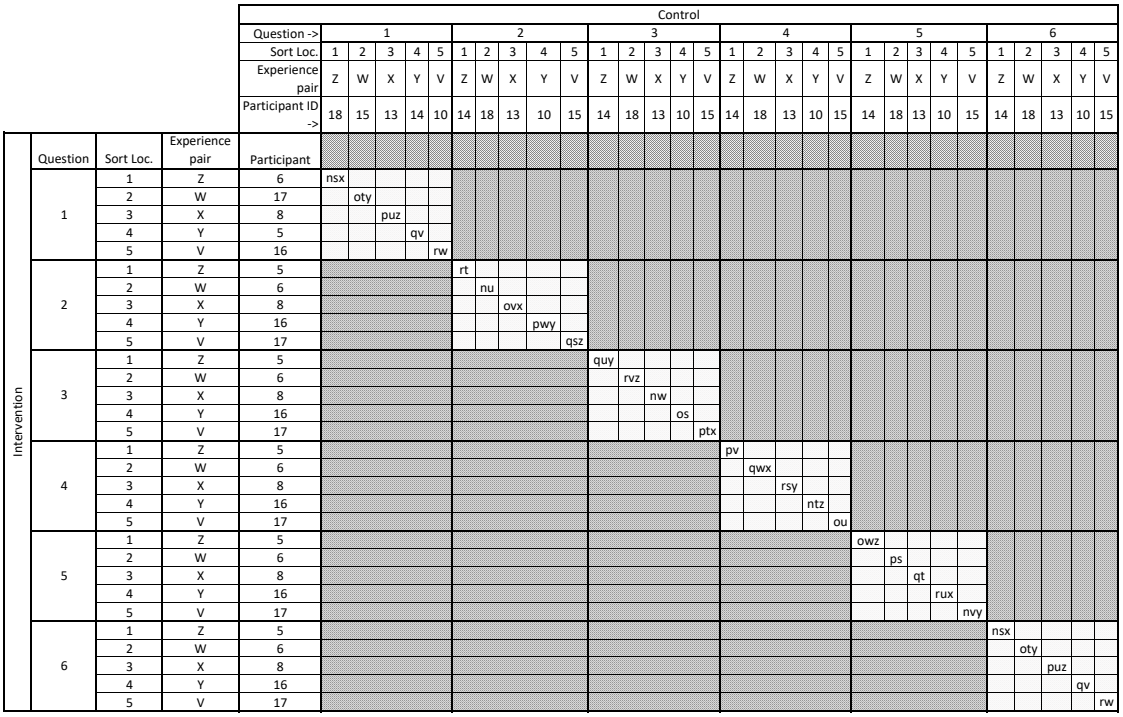
Figure 5.7: Screenshot showing the sort of Sub-phase 4.1 *experience pairs* (*V* to *Z*) into control or intervention cohorts.

Sub-phase 4.2 allocation Sub-phase 4.2 participants were randomly allocated to a position within a test plan (Figure 5.8), designated with a lower case letter. To mitigate unconscious bias, experience pairs were also randomly allocated their position in the test plan, indicated by a number 1 to 5 in the column and row labelled “Sort Loc.”. Figure 5.9 shows screen captures from the random allocation process using Algorithm 1[15], and a marginal total summary of the allocation for each experience pairing, indicating balanced scrutiny for each. Although cumbersome in set-up, sorting the participants in this manner provided a clear audit trail to aid administration of the experimental set-up, and to trace any issues.

Participant training Given that engineers are trained to assess problems objectively, it was expected that the intervention cohort may experience difficulty in working within a constructivist paradigm to make a decision about ostensibly physical phenomena. The specific application of an argument structure to a decision problem was also expected to be a new concept to all participants. The Sub-phase 4.1 intervention cohort received training in the novel framework in order to minimise the effect of these tensions and participate meaningfully in the experiment.

Training consisted of two aspects: written training guides (a short reference

[15]The sorting algorithm applied within Microsoft Excel necessitated an “instantaneous” selection to be made of the sort locations, because the random number function RAND continues to generate new random numbers, changing the allocation continually.



					Control																											
				Question ->	1					2					3					4					5					6				
				Sort Loc.	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
				Experience pair	Z	W	X	Y	V	Z	W	X	Y	V	Z	W	X	Y	V	Z	W	X	Y	V	Z	W	X	Y	V	Z	W	X	Y	V
				Participant ID ->	18	15	13	14	10	14	18	13	10	15	14	18	13	10	15	14	18	13	10	15	14	18	13	10	15	14	18	13	10	15
Intervention	Question	Sort Loc.	Experience pair	Participant																														
	1	1	Z	6	nsx																													
		2	W	17		oty																												
		3	X	8			puz																											
		4	Y	5				qv																										
		5	V	16					rw																									
	2	1	Z	5						rt																								
		2	W	6							nu																							
		3	X	8								ovx																						
		4	Y	16									pwy																					
		5	V	17										qsz																				
	3	1	Z	5											quy																			
		2	W	6												rvz																		
		3	X	8													nw																	
		4	Y	16														os																
		5	V	17															ptx															
	4	1	Z	5																pv														
		2	W	6																	qwx													
		3	X	8																		rsy												
		4	Y	16																			ntz											
		5	V	17																				ou										
	5	1	Z	5																					owz									
		2	W	6																						ps								
		3	X	8																							qt							
		4	Y	16																								rux						
		5	V	17																									nvy					
	6	1	Z	5																										nsx				
		2	W	6																											oty			
		3	X	8																												puz		
		4	Y	16																													qv	
		5	V	17																														rw

Figure 5.8: Allocation plan illustrating that each Sub-phase 4.2 participant (*n* to *z*) was allocated to each test case scenario and each experience pairing.

guide as an aid for reference during the experiment, and a longer more descriptive manual, Appendix L); and a one-to-one virtual meeting to ensure that intervention participants had correctly understood the concepts and had the opportunity to clarify any questions regarding use of the framework. Timing of the training was left to the discretion of participants within the overall deadline set for completion of their participation. The training commitment was expected to be 2 hours although this duration was not recorded.

Although the model was built upon foundations from argument and decision theory, complete training on essential skills from both fields such as statistical methods, heuristics and biases, and argument fallacies relevant to decision-making was not provided. This information was not deemed essential to test the functional aspects of the novel framework in Phase 4, and so only content pertaining to understanding the applied form of the novel framework was instructed upon. The content of the training packages is included at Appendix L.

The lack of independence between the lead researcher and the delivery of

***Experience Pairing* Sort for 4.2**

Pairing	[redacted]	Rand	Phase 4.2 Sort Location	Instantaneous Allocation	Selection (8 Nov 20, 1545)
Y		0.875	1	X	Z
Z		0.139	2	W	W
X		0.906	3	Z	X
V		0.954	4	Y	Y
W		0.479	5	V	V

(a) Random allocation of experience pairings to sort location, see Figure 5.8.

					Allocation of *experience pairing* to 4.2 participant					
Lettered 4.2 Test Allocation	Participant ID	Rand	Instaneous Allocation	Allocation (7 Nov 20, 1721)	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
n	1	0.104	*u*	*v*	Y	X	W	Z	V	Y
o	2	0.311	*s*	*n*	Z	W	X	Y	V	Z
p	3	0.559	*p*	*u*	X	W	Z	V	Y	X
q	4	0.565	*q*	*s*	Z	V	Y	X	W	Z
r	7	0.156	*v*	*p*	X	Y	V	Z	W	X
s	9	0.084	*x*	*y*	W	Y	Z	X	V	W
t	19	0.699	*n*	*q*	Y	V	Z	W	X	Y
u	11	0.259	*w*	*w*	V	Y	X	W	Z	V
v	12	0.017	*r*	*r*	V	Z	W	X	Y	V
w	20	0.303	*t*	*t*	W	Z	V	Y	X	W
x	21	0.461	*o*	*x*	Z	X	V	W	Y	Z
y	22	0.009	*z*	*z*	X	V	W	Y	Z	X
z	23	0.432	*r*	*o*	W	X	Y	V	Z	W

Pairing	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	*Pairing Marginal*
V	*2*	*3*	*3*	*2*	*3*	*2*	*15*
W	*3*	*2*	*3*	*3*	*2*	*3*	*16*
X	*3*	*3*	*2*	*3*	*2*	*3*	*16*
Y	*2*	*3*	*2*	*3*	*3*	*2*	*15*
Z	*3*	*2*	*3*	*2*	*3*	*3*	*16*
Case Marginal Total	*13*	*13*	*13*	*13*	*13*	*13*	

(b) Sub-phase 4.2 participant allocation to lower case letter identifier.

Figure 5.9: Allocation of participants and experience pairings in preparation for Sub-phase 4.2 testing.

training to participants is a limitation, but one that was insummountable within the bounds of the academic award. Examples discussed in person and provided in the training manuals did not refer to the test case context, and nor were they representative of the test case decision problems. One participant (Participant SIRM-Phase4-2020_17) proceeded with the test cases after only reading the training material, engaging with virtual training after the first case. Reflection on the quality of the response identified that this did not influence the “completeness” of the participant’s answers (Chapter 6).

5.2.3 Data collection

Owing to the experimental constraints discussed earlier, an online survey tool (Qualtrics, 2020) was utilised as the data collection method. The platform provided sufficient flexibility to present test cases, collect data for both intervention and control cohorts, and establish the individual comparisons from each Sub-phase 4.2 participant.

5.2.3.1 Pilot testing

Prior to the launch of Sub-phase 4.1, the novel framework, training and online elements of Sub-phases 4.1 and 4.2 were pilot trialled. The responses informed a number of immediate improvements:

- Novel framework design
    - Use of Toulmin argument lexicon was perceived by pilot tests as too abstract so the language was made more familiar for practitioners (see Section 5.1.3.3).
    - Training notes were not user focussed and read too academically. Content was adjusted to make each argument category clear and memorable through diagrams and use of colour.
    - A quick reference guide was produced with more relatable examples in order to help users adapt to the use of the tool quickly, and to enable a reference for participants during the experiment.
    - More examples were added to aid understanding of the application of the tool to practical situations.
- Sub-phase 4.2
    - The supplied glossary was misleading, resulting in presentational improvements.

5.2.3.2 Consent

All participants in Phase 4 were required to indicate their informed consent. Participants were provided with "Participant Information Sheets" relevant to their cohort, consent forms and Ministry of Defence Compensation information, which were a mandatory requirement set by MODREC. Since all data would be collected through Qualtrics, a pre-survey consent question was provided. Failure to provide consent would end participation, and no data would be collected. All participants provided their consent, indicated through their selection in Qualtrics.

5.2.3.3 Sub-phase 4.1 intervention cohort

The Qualtrics platform allowed a common survey to be accessed by all intervention participants. A summary of the questions asked for both Sub-phase 4.1 cohorts is at Appendix K.3.

After providing their consent, participants were asked to provide details of their experience (see Section 5.2.2.3) before being given access to the test case context information (see Appendix K). The participant could then advance onto the first case, the countdown timer for that case would begin, and the participant could view the case information from the link provided (Figure 5.10a). Each case required the participants to make a specific decision with relation to the case information presented (Appendix K).

Translation of the novel framework to Qualtrics required the form to be split into a mix of imagery prompts and selection buttons, in order to enable participants to use the novel framework as intended. Each of the Gates required a yes/no selection to be made (Figure 5.10a), before the "Argument" portion of the framework was displayed to allow participants to populate their information into the relevant elements (Figure 5.10b). Given the simulated time pressure, it was expected that participants would not populate more than 4 argument clauses for each test case decision (see Figure 5.11b and Annex A to Appendix L.1).

After each case, participants were asked to indicate how confident they felt presenting their judgement on the test case, and whether there were any aspects

of their judgement that they had difficulty communicating (Figure 5.11a).

On completion of the sixth (final) test case, participants were asked to reflect holistically on the novel methodology and its suitability for judgement-based risk decision-making in operational engineering environments.

Each completed case response was translated into the form envisioned for the novel framework (Figure 5.11b) and given a unique identifier, for example:

4.1i_5_2Y

Where $4.1i$ refers to Sub-phase 4.1 intervention, 5 refers to the participant ID number (n as identified in Section 5.2.2.4), 2 refers to the case number, and Y refers to the experience pair. Labelling each entry in this manner was essential for administrating the organisation of the Sub-phase 4.2 data collection, without losing configuration control and data efficacy.

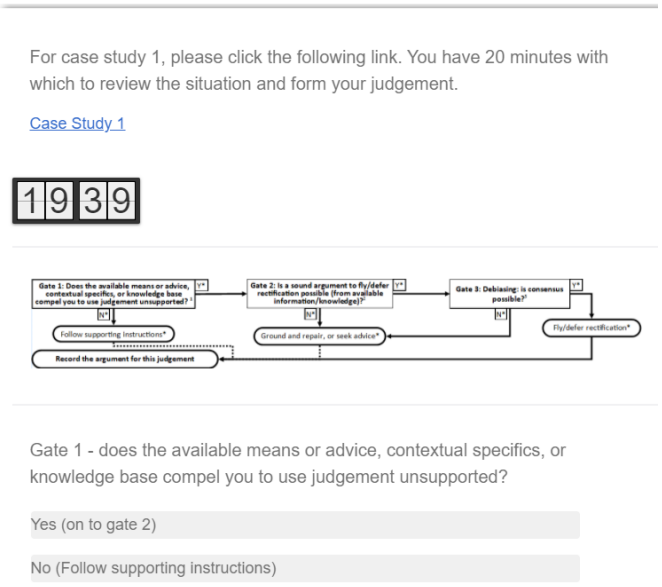
5.2.3.4 Sub-phase 4.1 control cohort

The process and case information provided for the control cohort mirrored the intervention cohort. The only difference was in the means with which the participants had to justify their decision. While the intervention cohort had a structured method, the control cohort had a simple open text box, unstructured to allow the participant to use the method they are most familiar or comfortable with (Figure 5.12a). As with the intervention cohort, Qualtrics allowed the same survey format to be accessed by all control participants.

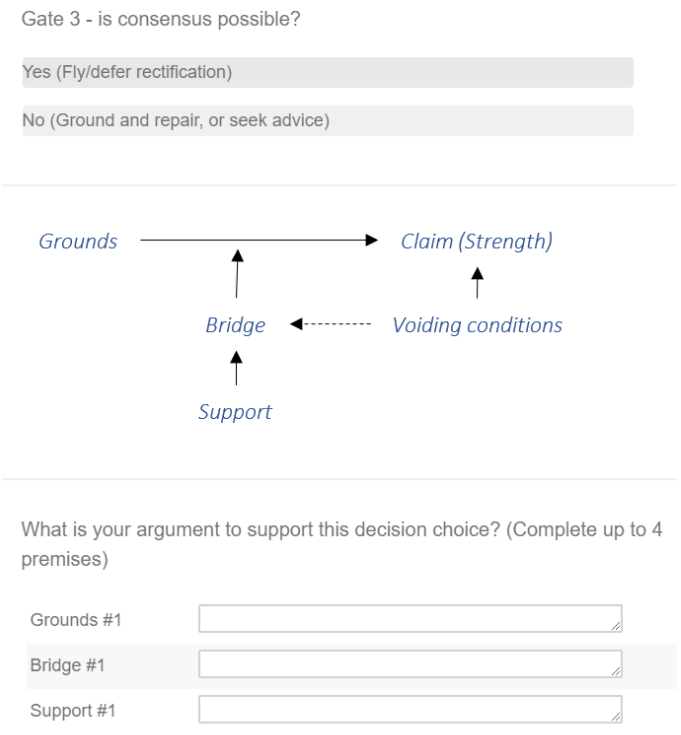
Each completed control cohort case response was translated on to a simple plain text document (Figure 5.12b) and given a unique identifier using the same file referencing convention as for the intervention responses.

5.2.3.5 Sub-phase 4.2 cohort

Irrespective of the obvious presentational differences between the intervention and control, the two different methods were referred to as Answer X (control) and Answer Y (intervention), to mitigate cognitive bias (Figure 5.13). Each participant

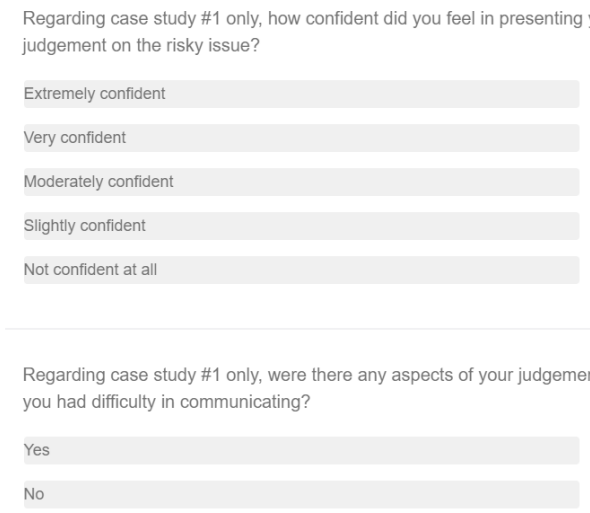


(a) Intervention participant screen capture during testing showing link to the case study question, the countdown timer, the heuristic decision tree prompt and Gate 1 question selection options.

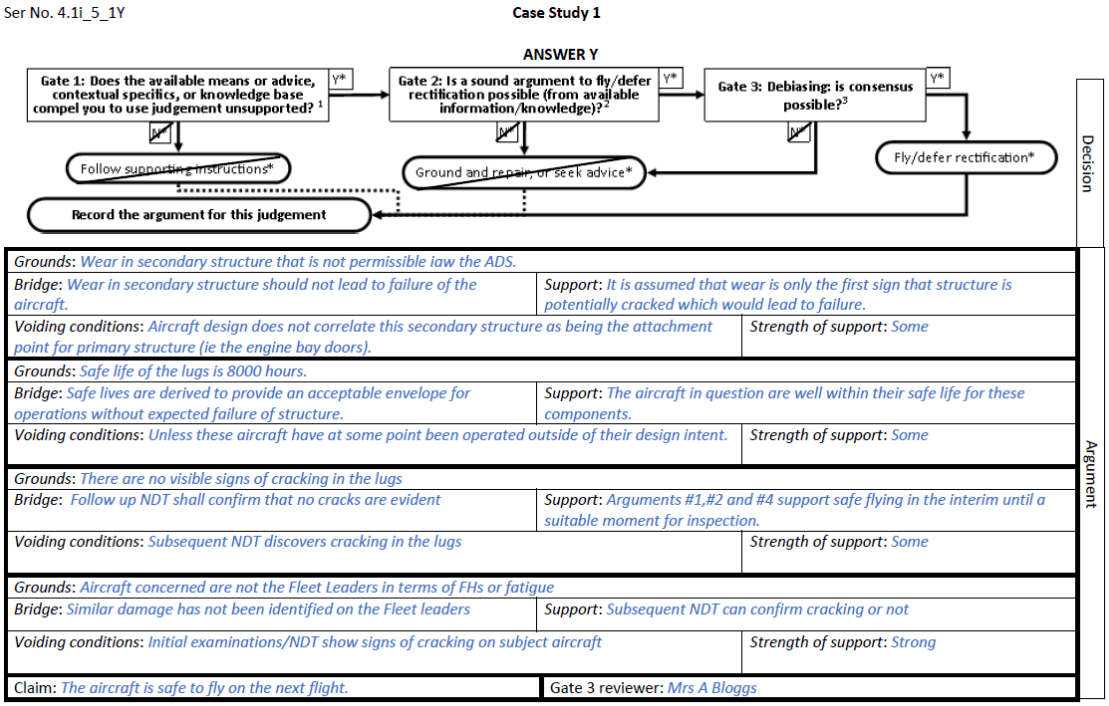


(b) Phase 4.1 intervention participant testing screen capture during testing showing Gate 3 question, a prompt of the Toulmin argument structure and the argument fields to capture participant responses.

Figure 5.10: Intervention participant data input fields, screen capture during testing.

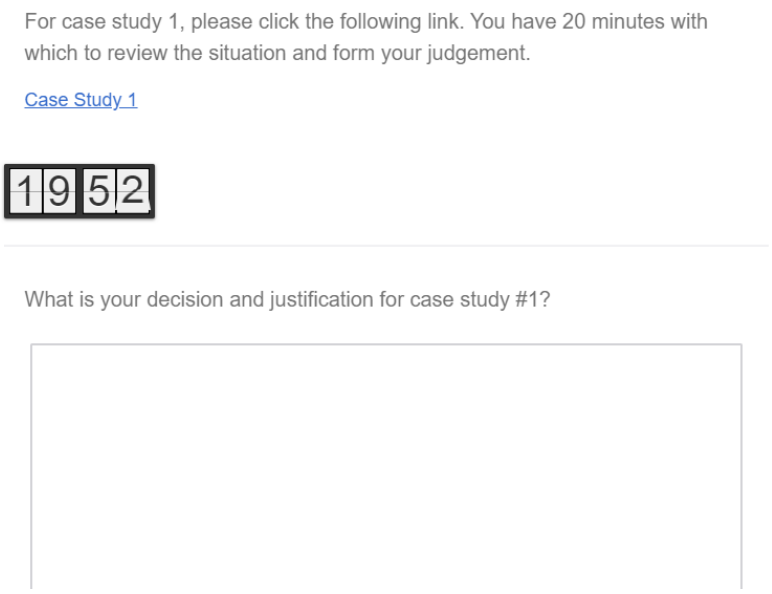


(a) Intervention participant post-scenario questions, screen capture during testing.



(b) Intervention participant response formatted as presented to phase 4.2 participants.

Figure 5.11: Phase 4.1 intervention participant testing screen captures.



(a) Phase 4.1 control participant screen capture during testing showing link to the case study question, the countdown timer and the free text field to capture participant responses.

Ser No. 4.1c_10_1V **Case Study 1**

ANSWER X

Both aircraft are mid-life and not the fleet leaders in usage (flying hours, landings, or fatigue usage), therefore these ac could be indicative of early wear. It has never been seen before though, and it is expected that the fleet leaders would have had the same inspection carried out. Therefore the exposure to the risk should be only since the last Depth maintenance.

The disposition of the fleet is unknown, therefore I do not know where the fleet leaders are based. This should be undertaken to ascertain where the fleet leaders are, and which ac are deployed where. Priority is airworthiness, but the impact on ops may help decision making if the ac are all recently out of depth (cycle duration dependant) Ultimately, more information is required.

Immediate visual inspection of the fleet will confirm the numbers of ac in this position. If it can be seen without removing the doors entirely. Integrity of the mounts is still there as long as there is no cracking, which needs to be checked on the depth ac, except that the hole is enlarged which may cause movement of the door. Inspection, if possible will provide further information on how many ac are affected. NDT of the ac in Depth should be carried out to detect any possible cracks and address the issue on these 2 ac.

If the mounting brackets were primary structure, then this should have been stated. Although this is never a reliable fallback and the assumption has to be made that this is just as important. Urgent request for OEM repair should therefore be actioned regardless.

(b) Phase 4.1 control participant response formatted as presented to phase 4.2 participants.

Figure 5.12: Phase 4.1 control participant testing screen captures.

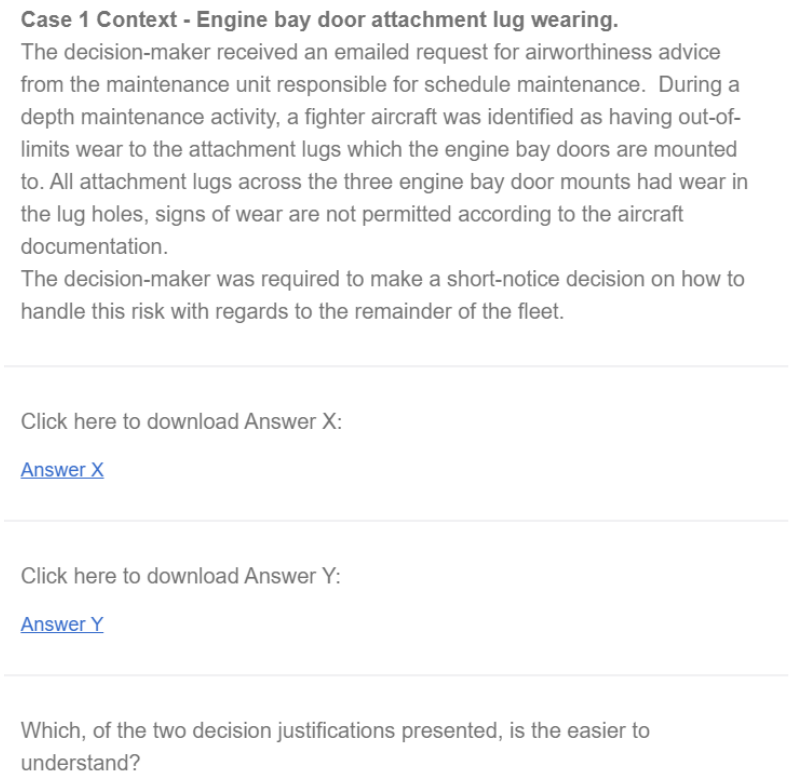


Figure 5.13: Phase 4.2 participant testing screen capture showing scenario context and options to view Answer X and Answer Y.

required access to their own personalised survey, allocated as illustrated in Figure 5.9b. Participants were provided with less contextual and scenario information than the Sub-phase 4.1 participants, enough to understand the decision justifications but not enough to form their own diagnosis of the situation, which could have prejudiced the review task through cognitive bias (Appendix K). Participants were then asked which answer was easier to understand, which gave greater transparency of the decision basis, and which provided a better justification of the decision (Appendix K.3).

5.2.3.6 Problems

The requirement to manually translate decision justifications from Sub-phase 4.1 into the prescribed formats for presentation to Sub-phase 4.2 participants resulted in some problems. The use of unique identifying answer numbers helped maintain control, however, a transcription error was still reported by a Sub-phase

4.2 participant within the Case 4 intervention answer from Participant 8. The time afforded to compile the data from Sub-phase 4.1 and prepare it for presentation to Sub-phase 4.2 was, in hindsight, insufficient despite having taken proactive measures to minimise error. Tracing the error, two other participants received the error-containing decision justification[16]. The participants were contacted post hoc to confirm whether they recalled seeing the error, and whether its presence had bearing upon their responses, with neither claiming that the error influenced their responses, so no further mitigating action was taken.

5.3 Chapter summary

The process followed in the development and testing of a novel framework has been described. Design of the framework was guided by a set of requirements extracted from the survey and case study analyses, resulting in a hybrid protocol that targets naturalistic decision-making. The framework guides decision-makers using an heuristic tree with binary choice decision gates, and supports justification using an argument schema to structure information used in decision-making.

The experimental strategy tests usability with real-world aircraft maintenance practitioners through realistic but fictional representative textual scenarios, which are completed under a soft time constraint. The decision responses formed using the novel framework are then compared with existing decision-making approaches. Pilot testing enabled the experimental strategy to be improved iteratively.

[16] Error tracing was possible as a result of the allocation plan and audit trail discussed in Section 5.2.2.4).

Chapter 6

Novel framework test results and discussion

As described in Chapter 5, testing of the novel framework involved a two-phase experiment. For the first phase, six representative and realistic but fictional scenarios were created, each concerning a typical aviation maintenance situation with a structural fault. Ten experienced practitioners were split into five pairs, which were assigned keeping experience levels similar within each pair. One member of each pair was then randomly assigned to the intervention cohort and used the novel framework for decision-making in the test scenarios. Intervention cohort members received training in the use of the novel framework. The other member of the pair was assigned to the control cohort, using their existing decision approach for the test scenarios.

In the second phase, 11 experienced practitioners were required to compare the decision justifications made by the intervention and control participants in the first phase. Each participant made six comparisons (one comparison for each scenario) and was randomly allocated a pair of decision justifications, seeing a justification from each pair at least once. A total of 66 individual comparisons were made by the 11 raters.

6.1 Results from the first phase of the safety decision experiment

Participants assigned to the intervention cohort underwent training to provide familiarisation on the framework design and example scenarios. The application of the framework to the test cases was considered a *structured* approach. One-to-one training sessions provided participants the opportunity to clarify their understanding, although one participant (SIRM-Phase4-2020_17) proceeded to the task, attempting test cases 1 and 2 before pausing their participation to engage in the training. All test cases were attempted by the intervention cohort, one participant (SIRM-Phase4-2020_16) chose not to provide any decision justification for test case 1, in the mistaken belief that it was not required if the decision chosen was to *ground* the aircraft at Gate 1. Use of the framework was generally consistent, though some individual participants (notably SIRM-Phase4-2020_6) had difficulty in applying the argument structure to the test cases.

Participants assigned to the control cohort required no training, since they were expected to justify their decisions using their own familiar approach. This was considered as an *unstructured* approach to the test case decision problems. All test cases were attempted by the control cohort, although the level of abstraction applied by each participant varied.

Results were considered against potential explanatory factors concerning participant experience[1], last judgement use, case number, and word count. Additional factors were identified during analysis, including completeness (how fully the new framework was applied) and decision alternative chosen (*ground* or *fly*).

[1] It was expected that a more experienced participant would provide a more informed justification, which may dominate a far less experienced participant. To prevent imbalance between control and intervention comparisons, participant experience in years was requested prior to the commencement of testing to allow allocation to experience pairs of similarly experienced participants, see Section 5.2.2.4.

6.1.1 Analysis of responses from the first phase of the safety decision experiment

The main effect to be measured from the first stage of testing was usability of the intervention method and support for decision-makers deliberating aircraft maintenance structural integrity faults and failures under time, resource and information constraints. Thematic analysis of participant reflective feedback was combined with self-assessed confidence, completeness of the decision responses, and an analysis of the information used in decision-making, to draw conclusions. The results presented in this section are a mixture of qualitative and descriptive quantitative results. They show that decision-makers were supported by the novel framework, and perceived the method to be usable under real-world operational conditions.

6.1.1.1 Framework feedback

Reflective responses from intervention participants indicated, without exception, that participants “agree” (median) that the method would be usable for handling aviation maintenance structural faults and failures in operational situations (Figure 6.1). Feedback regarding the usefulness of the intervention commended the method for its complementing of cognitive function, encouraging personal reflection on the judgements made, and being able to demonstrate and justify the decision basis:

– Participant SIRM-Phase4-2020_16 - *“It led me through a structured thought process which enabled me to articulate my reasoning when analysing risk. It also made me consider and express the ’so what’ of decisions.”*

– Participant SIRM-Phase4-2020_17 - *“The process of decision-making was logical, and followed instinctive thought patterns and progressions, considering key factors to the eventual decision.”*

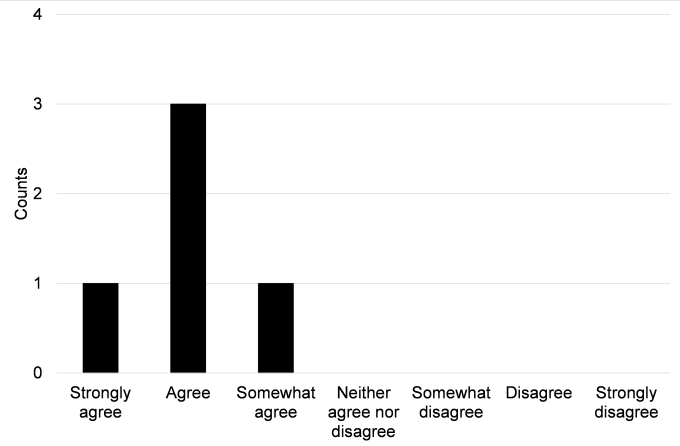


Figure 6.1: Proportions of responses by intervention cohort participants to experiment reflection question: “How far do you agree with this statement: this method I applied for justifying my judgement would be usable in an operational engineering environment to support judgement-based risk decision-making?”

Criticisms of the approach related to understanding of the Toulmin argument schema, which was expected after only a brief training session. However, other difficulties were common with control participants, namely how to verbalise intrinsic or intuitive perceptions of a situation. Figure 6.2 demonstrates this for one participant, who justified a decision to *ground* the aircraft, but could not decompose their justification any further than as a result of their “risk appetite”. Thematic analysis of the feedback provided resulted in the thematic map in Figure 6.3, showing the full list of themes extracted.

While the reflective responses from the intervention cohort provide insight into perceptions about the usability of the new method, control participant responses provide a measure of self-belief. The control cohort were left to select and apply their own approach as they saw appropriate for the situation, and they too “agree” that their method would be usable for handling structural integrity faults in operational situations. This is indicative of control participant individual belief in

Grounds: *primary structure which transfers load between the port wing and the fuselage - low risk appetite*	
Bridge: *personal risk appetite*	**Support:** *primary structure involved in transfer of critical load in key area of structure*
Voiding conditions: *if generic repair advice applies (Vol. 6 equivalent) to this structure or if ADS lists acceptable damage criteria.*	**Strength of support:** *Strong*

Figure 6.2: Participant SIRM-Phase4-2020_16 response to case 4.

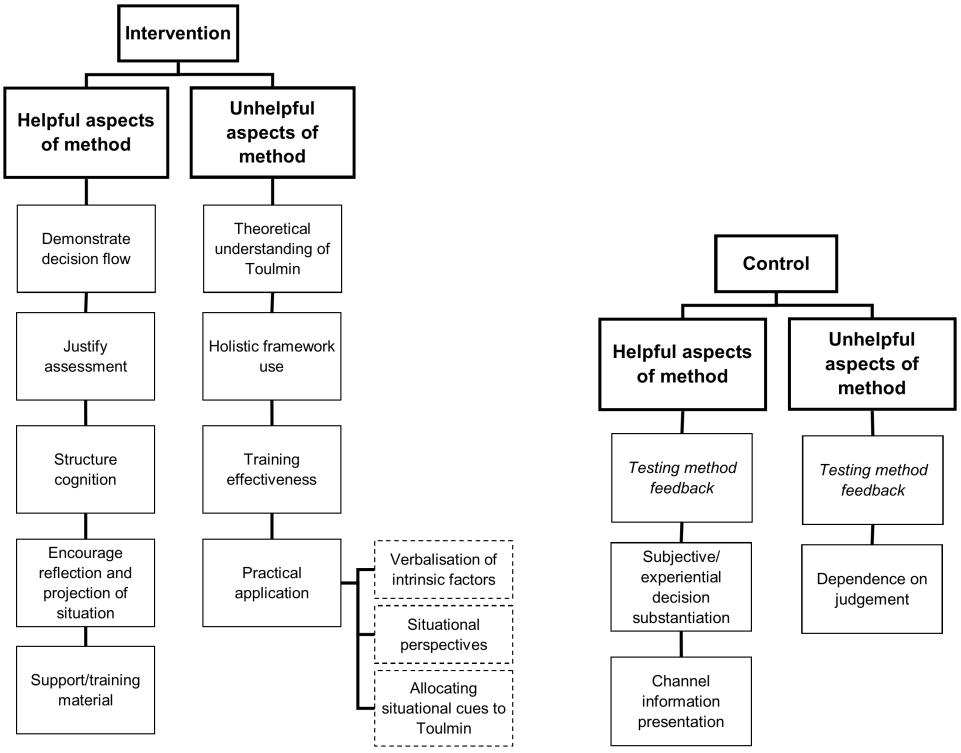


Figure 6.3: Phase 4 intervention and control cohort framework feedback.

their own approach, which is moderated by their own experiences and knowledge. Control participants, it seems, have no reason to doubt their own approach. Insignificant correlations between operational suitability and confidence[2] suggest that confidence and perceptions on suitability may be coupled, but a larger sample is required to improve statistical strength.

Following each test case, participants were asked to detail any aspects they had difficulty communicating. These responses were thematically analysed, which highlight the difficulties participants had with externalising their internal processes:

– *Communicating internal processes.* Metacognition; verbalising intrinsic assessments.

– *Deliberative activities.* Decomposing the problem; identifying the usefulness of information; indicating secondary activities beyond the decision alternative chosen.

[2]Control ($R = 0.518, p = 0.371$; intervention ($R = 0.645, p = 0.239$).

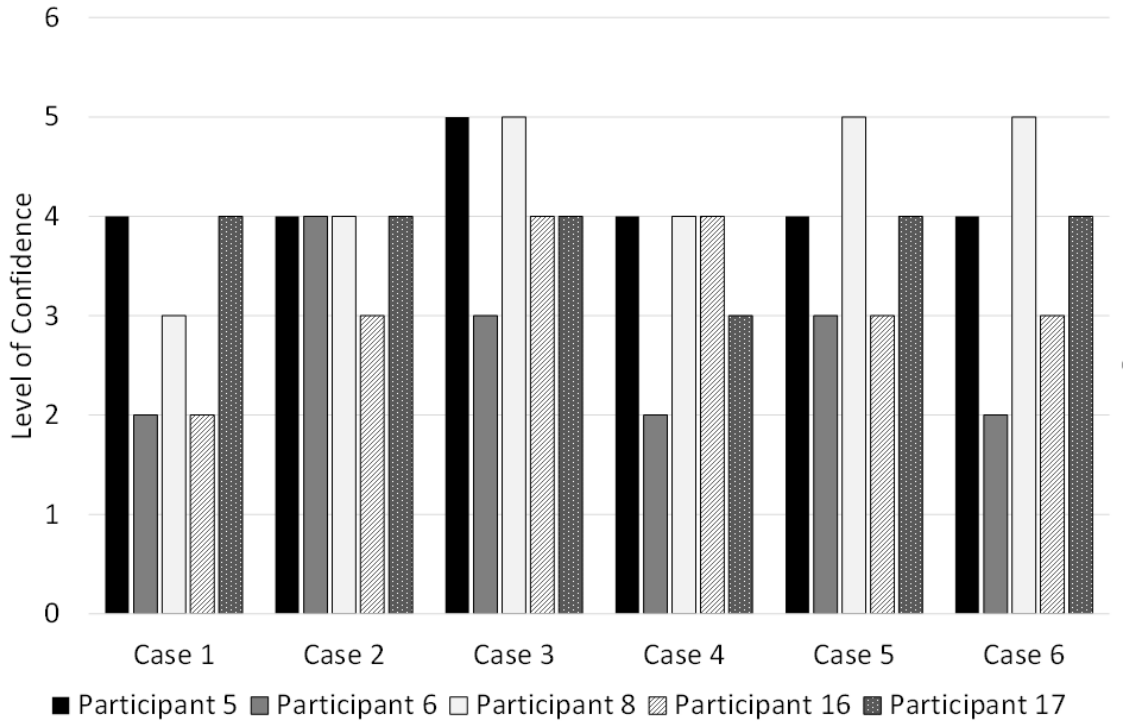


Figure 6.4: Intervention cohort confidence levels for each test case.

– *Experimental practicalities.* Framework use; training and experiment; time pressure restricting justification.

Similarly, control participant difficulties overlapped the intervention feedback with regards internal processing:

– *Internalised processing.* Subjective/instinct about scenario; causal visualisation; management considerations - action necessary (mitigation, intervention, applicability).

– *Information dependence for decision and action.*

6.1.1.2 Confidence

Median intervention participant confidence was “very confident” and increased on average through the test. Average range was 1.6[3] and the standard deviation for the mean confidence levels of all participants, $\sigma_{mean} = 0.640$, was larger than the

[3]A Likert scale for confidence was used which provides a usable measure for comparison.

control, which had a range of 1 and a standard deviation $\sigma_{mean} = 0.388$. This is expected given the less familiar approach of the intervention cohort, and control participants being consistently self-assured in their own abilities. However, the difference is not statistically significant at the 5% level. The lowest and second lowest intervention cohort median confidence (across all cases) was seen on test case 1 (moderately confident) and test case 4 (very confident) respectively. Figure 6.4 shows the confidence levels across all test cases for the intervention cohort.

6.1.1.3 Completeness

Another measure of the degree to which decision-makers were supported by the intervention was by considering how complete the decision responses were. This measure indicated the extent to which participants complied with the Toulmin schema and was also considered as an explanatory factor for comparisons observed in the second stage of the framework testing experiment. An ordinal scale was created to indicate the extent to which a participant's response complied (functionally) with the Toulmin argument form:

- *Complete. Requisite elements of Toulmin structure filled in. They may not be logically valid, but the elements of the structure are complete.* This was a compromise, given the relative novice status of the participants in applying the Toulmin schema. Figure 6.5a illustrates an extract from one response in which the *support* lacks general support for the *bridge*, instead inferring through evidence rather than accepted principles, committing "post hoc ergo propter hoc" a fallacy of "after the fact, therefore because of the fact". In addition, the *voiding conditions* are not clearly explained.
- *Partially Complete. An argument is identifiable but there are some elements of the Toulmin structure missing.* Figure 6.5b gives an extract from a response. One clause is complete, but the second clause highlights a *support* statement that does not underpin the *bridge*, and the *voiding conditions* are not completed. However, the information recorded is supportive of the *claim*.

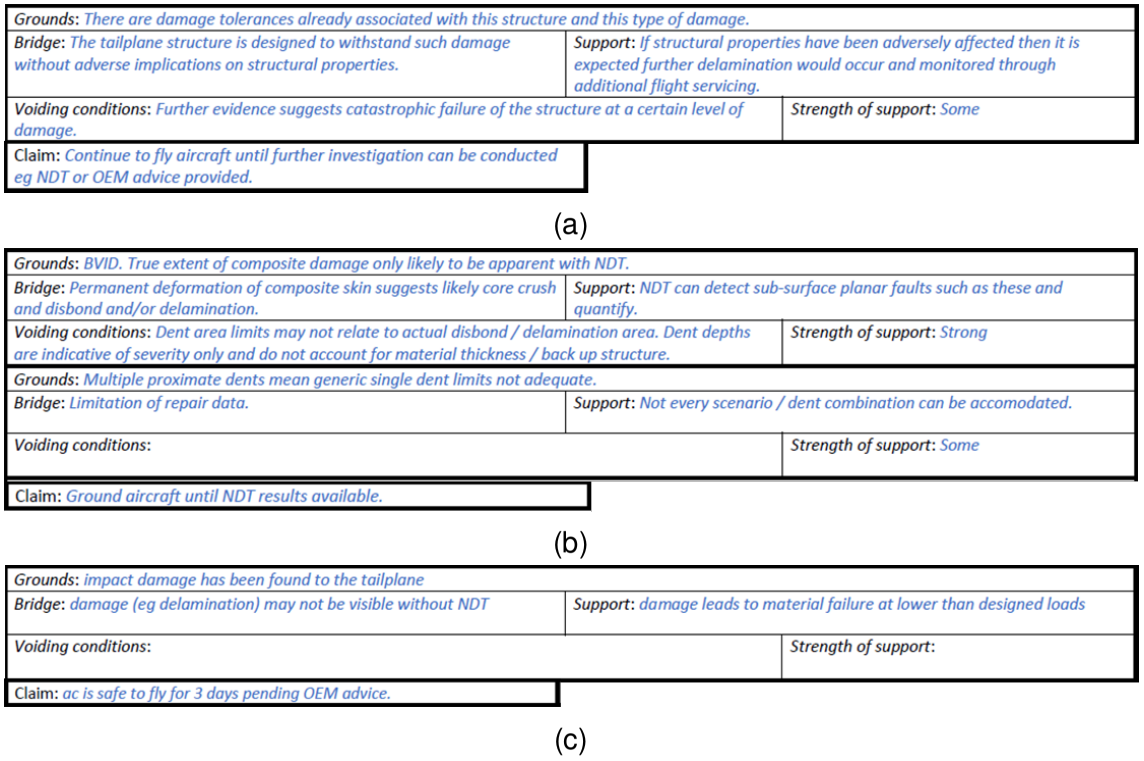


Figure 6.5: Examples of the varying degrees of completeness submitted by Sub-phase 4.1 intervention participants: 6.5a is an extract from a *complete* response; 6.5b is an extract from a *partially complete* response; and 6.5c an extract from an *incomplete* response.

– *Incomplete. Incorrect application of the Toulmin structure applied, and no clear argument is presented.* Figure 6.5c gives an extract from one response, which shows a clause that does not support the overall *claim*, and is missing *voiding conditions* and *strength of support*.

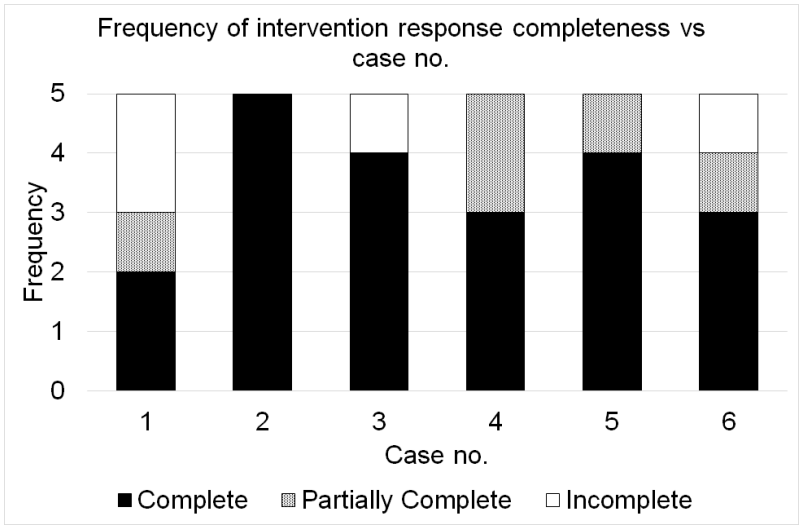
All intervention responses were subjectively assessed against this scale and 4 were categorised as "incomplete", 5 were categorised as "partially complete", and 21 were considered "complete". Figure 6.6a shows the profile of decision justification completeness across the cases, while Figure 6.6b shows the completeness by participant. A significant association was identified between percentage "complete" and participant last judgement use suggesting that those who made such judgements more recently had higher completeness levels ($R = -0.953, p = 0.01$). This is shown in Figure 6.6c. A high but statistically insignificant correlation was also noted between percentage "complete" and participant experience (years) ($R = -0.815, p = 0.093$). This suggests that recency has bearing on individual ability to use the functions in the intervention model, namely, argument. More evidence is necessary to conclude that less experienced participants may be more agile to using new approaches, or alternatively, that more experienced participants are more disagreeable with new approaches.

6.1.2 Further analysis of the first phase responses

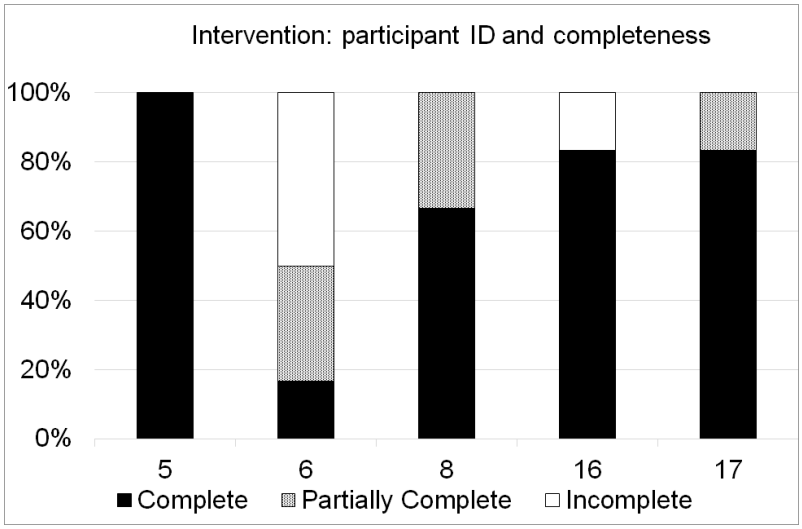
Analysis was also undertaken of test case completion times, alternate choice selection, clarity of decision justifications, and differences noted between cohorts in the information used in the decision justification.

6.1.2.1 Information use

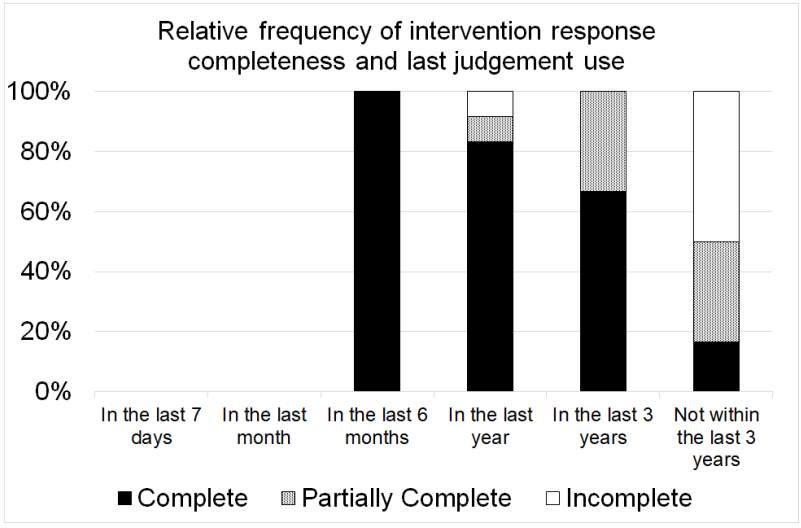
Measuring the quantity and quality of information used by participants was an opportunistic analysis that was not originally considered as a measure of support for decision-makers using the intervention method. By counting the number of case lines that were referenced in decision justifications, it was found that intervention participants referenced significantly more lines of the



(a)



(b)



(c)

Figure 6.6: Descriptive statistics of completeness variation across intervention participants.

	Control			Intervention		
	N	σ	95% CI	N	σ	95% CI
Fly	9.31	3.79	$\pm$2.69	13.5	3.91	$\pm$2.34
Ground	9.29	4.24	$\pm$2.54	13.3	6.26	$\pm$4.45
Overall	9.3	3.98	$\pm$1.72	13.4	4.97	$\pm$2.14

Table 6.1: Comparison of control and intervention cohorts information use, measured by counting each line of the given test cases within participant responses. N is the average number of counts across cohort participants.

test cases compared to control responses (two-tailed Mann-Whitney U Test, $W = 686.5, p < 0.001$). Although this does not offer direct evidence of improved support for decision-makers, it is of interest because it highlights how participant behaviour and handling of the given information differed. Table 6.1 provides the comparison between control and intervention cohorts. A statistically significant and high correlation was identified between control participant mean case line count and participant last judgement use ($R = 0.994, p < 0.001$), and a high, but statistically insignificant, negative correlation between control participant mean case line use and participant experience (years) ($R = -0.802, p = 0.102$). This finding suggests that the intervention cohort was influenced less by recency of judgement use in their use of available information.

Measuring the information used was achieved by splitting the test case narratives given to the participants into individual information segments, and then carrying out a content analysis of all responses to count when a particular information segment was referred to. The case lines in the test scenarios were segregated so that each line contained one actionable cue. For example, in test case 2, one section of the test case narrative read:

> *"There are no indications of consequential damage to the flap or mating wing structure, such as a fretting or wear, and there have been no recent airborne indications of flap failures attributable to retaining bolt failures. There are 2 actuation attachment points per flap, and one flap on each wing, meaning there are 4 similar bolts across the flap system on the aircraft."*

This was divided and identified as the following lines:

- Lines 11/12. *There are no indications of consequential damage to the flap or mating wing structure, such as a fretting or wear,*
- Line 13. *and there have been no recent airborne indications of flap failures attributable to retaining bolt failures.*
- Line 14. *There are 2 actuation attachment points per flap,*
- Line 15. *and one flap on each wing,*
- Line 16. *meaning there are 4 similar bolts across the flap system on the aircraft.*

Where respondent answers directly or indirectly referred to a particular line, a count would be tallied.

Information used for decision-making In addition to the quantitative comparison detailed above, qualitative analysis was carried out to identify the most common information *class* used, differences between control and intervention cohorts in the information *class* most relied upon, and how the resulting decisions used the information. Classes of information were derived and labelled from survey result thematic maps. Question 19[4] (Table 6.2) identified information considered necessary to make a decision, which were called the *case information classes*. Question 20[5] (Table 6.3) identified activities, behaviours and considerations that were part of the decision-making process, which were called *decision classes*.

Common case lines were identified as those with a median count (across each of the intervention or control cohorts) of at least 1. Using a content analysis, the wording of those common case lines were then codified using the *case information classes*. The corresponding decision responses that used the common case lines were then codified using the *decision classes*.

[4]Survey Question 19: “What information is essential in order to make a decision upon an in-service structural integrity risk?”

[5]Survey Question 20: “Describe the decision-making process regarding in-service structural integrity risks.”

Information Case Class		
Environmental Factors	Influence of operating environment on practical/ methodological decision aspects	e1
	Holistic operational necessity/need	e2
	Perceived time available	e3
Practical solution constraints	Published limits for component at risk	p1
	Maintenance considerations	p2
	Mitigation options	p3
Measurable information	Fault and relevant maintenance history	m1
	Damage/fault information	m2
	Usage data	m3
	Related fleet usage data	m4
	Privileged design data	m5
Individualistic	Awareness of information state	i1
	Individual judgement threshold	i2
	Need for advice	i3
	Human error considerations	i4
Risk aspects	Identification	r1
	Consequence focussed	r2
	Risk factor	r3

Table 6.2: Case information class codes, extracted from survey Question 19 thematic analysis results.

Decision Class		
Risk determination	Identify a risk	R1
	Calculate/determine risk	R2
	Classify risk	R3
	Apply risk classification	R4
People dependence	Identify other people for consensus	H1
	Include stakeholder groups	H2
	Identify other people in decision-making	H3
Performance considerations	Assess localised performance impact	P1
	Identify influence of operational requirement	P2
	Identify intervention impact on organisational output	P3
Mitigation factors	Create mitigation options/decision alternatives	M1
	Identify actions required by mitigation options	M2
	Identify temporal/time factors	M3
	Identify resources necessary for mitigation	M4
Fault information	Identify related occurrence data	I1
	Identify dependence upon procedures and processes	I2
	Treating situation analytically (use of approved technical information)	I31
	Treating situation analytically (visualising or determining sequence of events)	I32

Table 6.3: Decision class codes, extracted from survey Question 20 thematic analysis results.

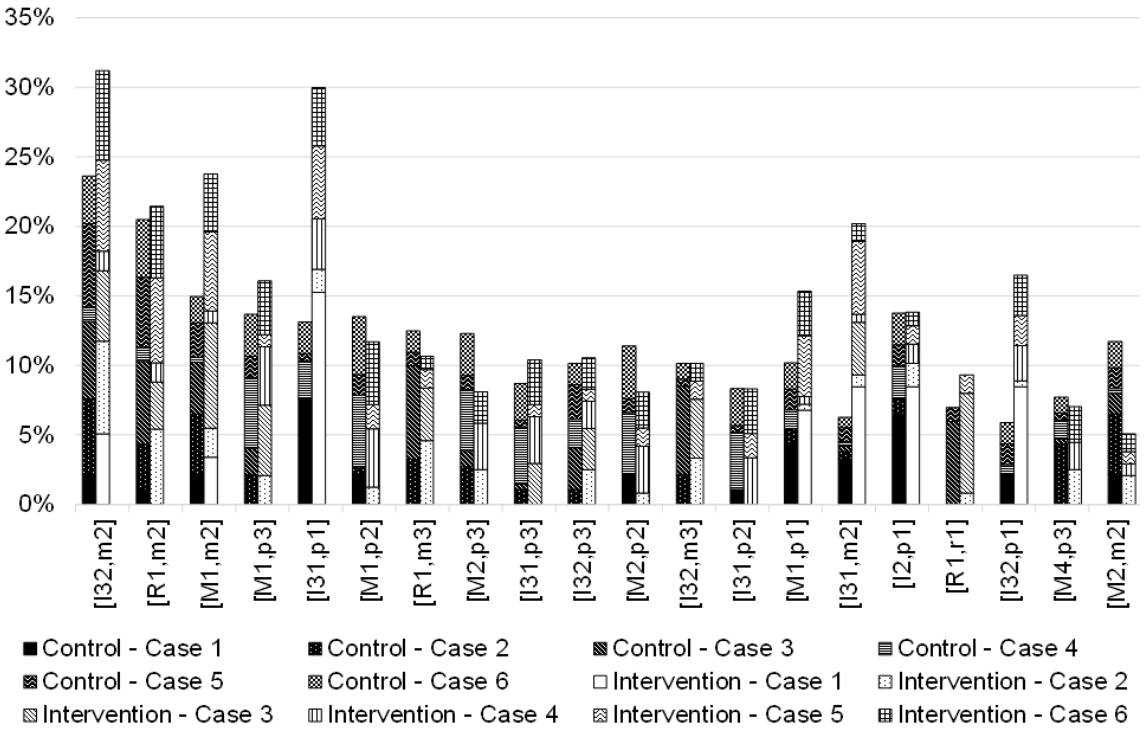
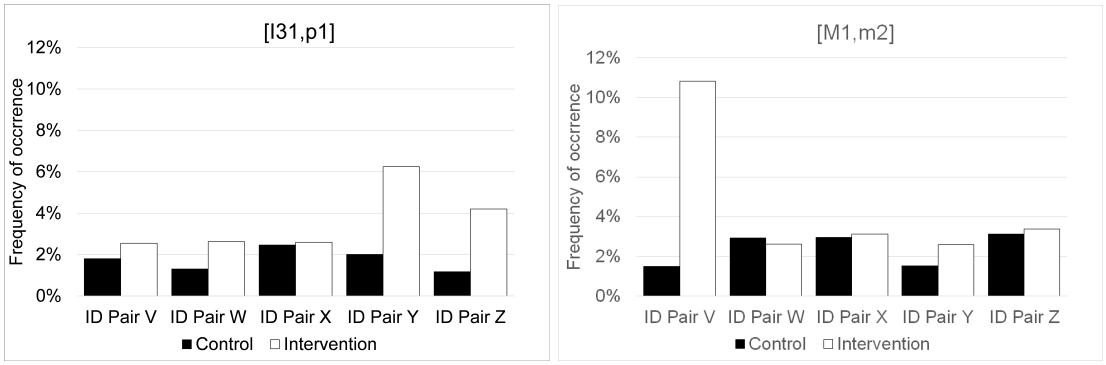


Figure 6.7: Comparison of the top 20 *class pairings* between control and intervention cohorts, with breakdown by individual cases.

This analysis produced a set of case information class codes paired to decision class codes for all control and intervention responses that used the common case lines. These are referred to as *class pairings*. Figure 6.7 shows the top 20 most frequently occurring *class pairings*, with the y-axis set as a relative frequency since presenting counts would show dominance by the intervention as they tended to use more case lines than the control responses. Decision classes are shown as upper-case, case information class as lower-case. Gross differences in the relative frequencies were identified between the control and intervention among the *class pairings* [I31, p1], [I31, m2] and [I32, p1]. The influence of *experience pair* was checked, which showed the same general trend, although individual results indicate the propensity for some participants to favour particular *class pairings*, as indicated in Figures 6.8a and 6.8b. Statistical analysis would account for confounding factors, however this was an opportunistic analysis to investigate potential effects within the available data set so an approximate analysis was deemed sufficient.



(a) (b)

Figure 6.8: Comparison of information case/decision class pairs, marginalised by *experience pairing*.

Common case information reliance In order to identify the most common case information class used, every case line referenced across all participants was translated into a count of the respective information case classes. Total counts were normalised against the quantity of each information case class given in each test case, to give a standardised measure of importance, and plotted as a relative frequency across all cases in Figure 6.9. A relative frequency is chosen to display the data in order to account for any variation in individual cases and across cohorts. The plot suggests that there are differences in the test case information use between the intervention and control participants.

These data were marginalised further to highlight decision differences (*fly* or *ground*), shown in Figure 6.10. The set of *discriminating* case lines were identified as disjoint case lines that were common between respondents who selected “fly” or “ground” but were not common in the conjunctive set (i.e. present in both sets of responses). This corresponds with the following condition:

$$D_{k,f} : \big(med(LineCount_{f,i}) \geq 1\big) \cap \big(med(LineCount_{g,i}) < 1\big)$$

$$D_{k,g} : \big(med(LineCount_{g,i}) \geq 1\big) \cap \big(med(LineCount_{f,i}) < 1\big)$$

Where D is the set of discriminating lines in case k, f and g indicate the counts of lines from *fly* or *ground* decisions respectively, and $i = 1...n$ indicates a given line number for a test case containing n lines. Discriminating lines were

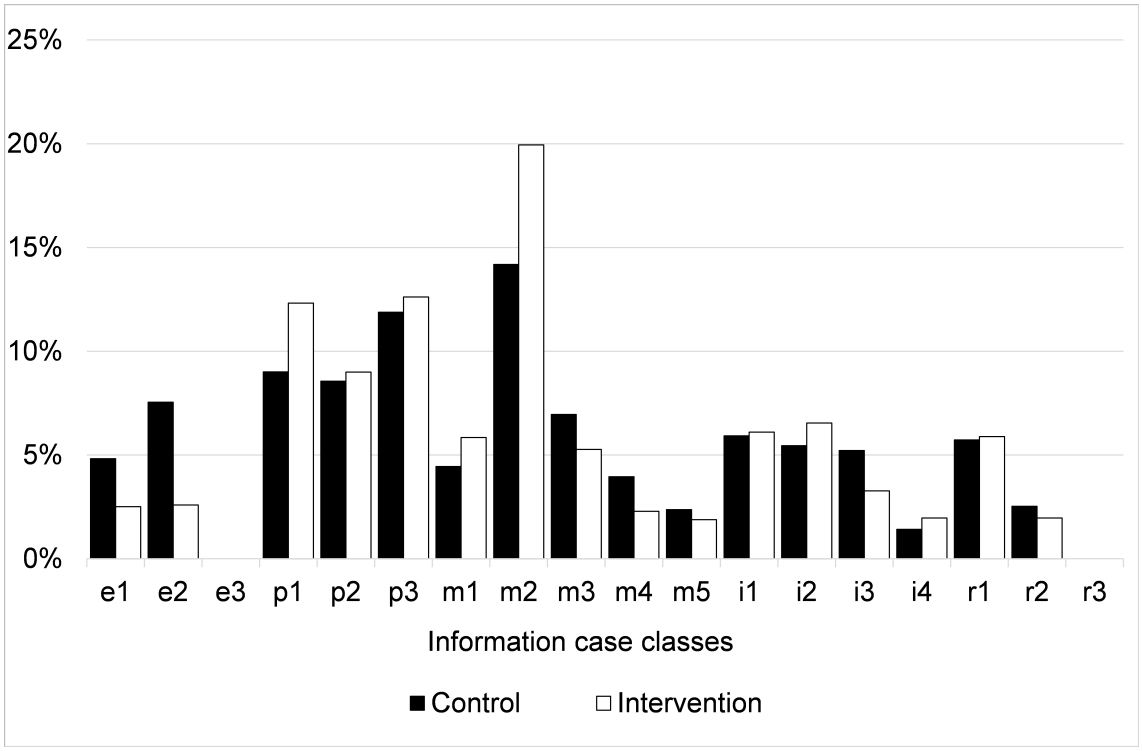


Figure 6.9: Comparison of normalised relative frequency (all cases) of case information classes used by participants.

translated into corresponding *case information class* and the counts of the classes were normalised. The resulting plot is given in Figure 6.10. *Gross differences* between classes were again used to discriminate between *fly* and *ground* choices. Differences across control and intervention participant responses, indicated in Table 6.4, suggests differences in what information is used to deliver a judgement to *ground* or *fly*.

6.1.2.2 Completion time

The soft time limit imposed on the each test was intended to simulate real-world pressure on participants to make a decision. Of the 60 decision responses across both control and intervention, there were 14 violations of the limit. 12 exceeded the limit by a maximum of 15% (3 minutes), 2 were gross violations from the same participant (SIRM-Phase4-2020_17) on cases 1 and 2 that exceeded the time limit by over 185%. Discussion with this participant identified that this resulted from their

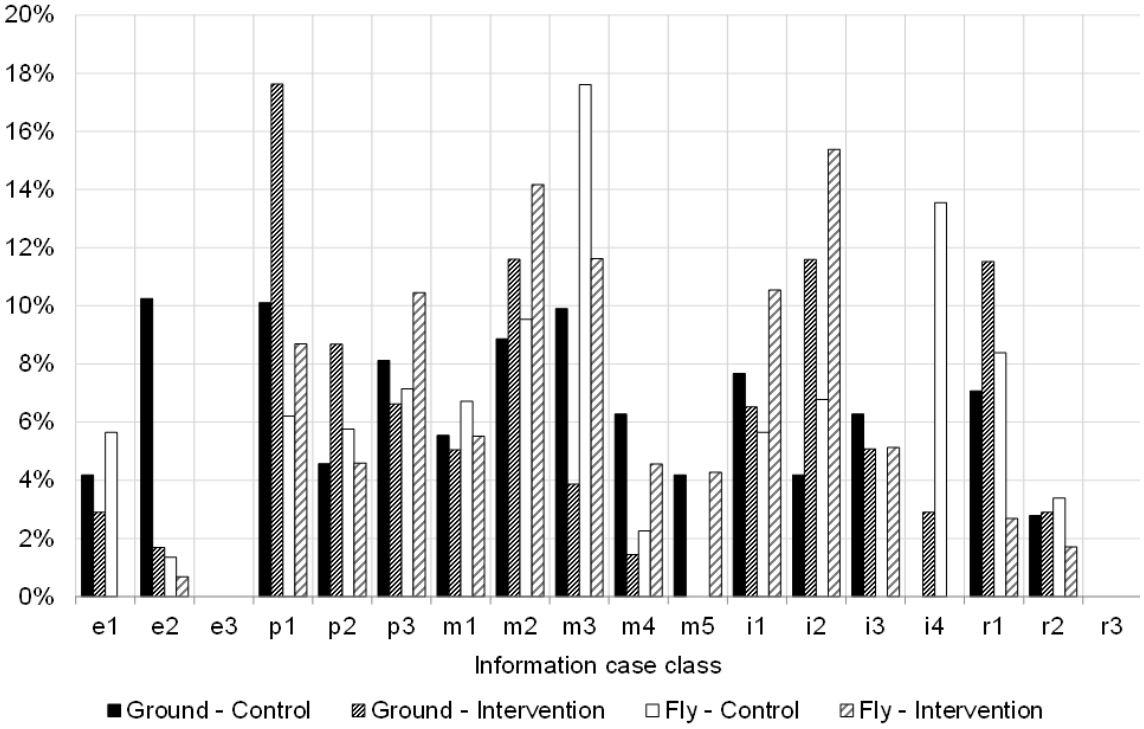


Figure 6.10: Comparison of normalised differences (relative frequencies) between case information classes, averaged across all cases for case lines identified as “discriminating”.

	Intervention		Control	
	Ground	Fly	Ground	Fly
Gross difference	e1,p1,**i4**, r1	m3,**m4**,**m5**	e2,**m4**,**m5**, i3	**i4***
No difference	p2, p3, m1, m2, i1, i2, i3,r2		e1,p1, p2, p3, m1, m2, m3, i1, i2, r1,r2	
No data	e3, r3			

Table 6.4: Differences in case information classes across control and intervention. underline indicates classes common to intervention and control responses; **bold** indicates opposite effects in intervention and control responses; *i4 was only present in one case (case 5).

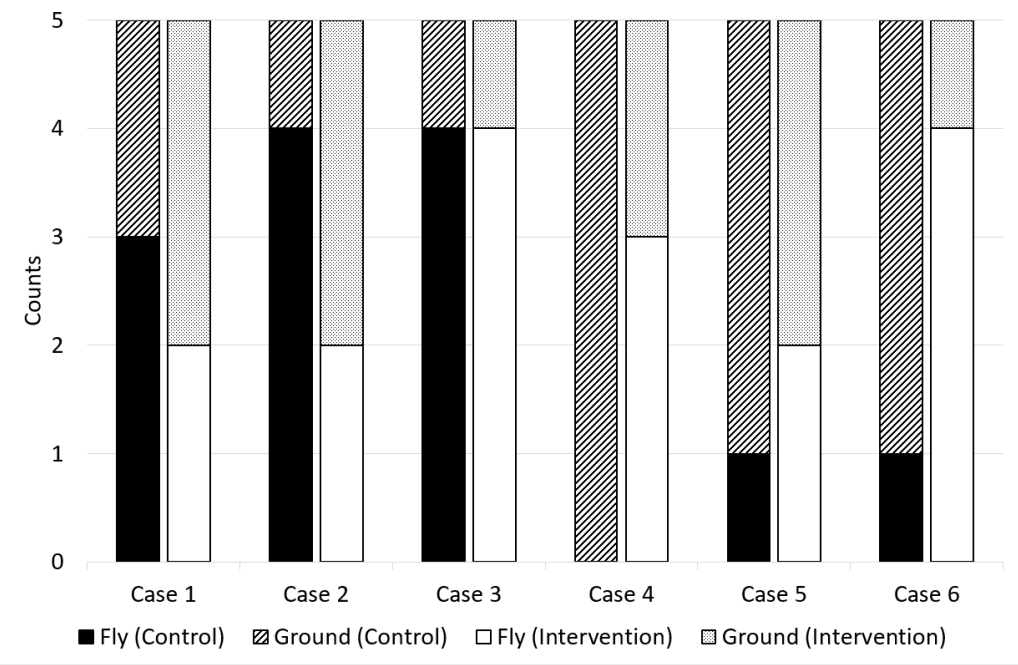


Figure 6.11: Comparison of control and intervention cohort decisions across each test case.

commencing the task without having completed the full training on the intervention method. Their responses had little variation in completeness or quality changes across the response set. No violations of the soft time limit were removed from the sample set, owing to the small sample set and Participant 17 considerations specifically. SIRM-Phase4-2020_16 had an outlying completion time owing to their premature response to case 1. As a result of the outliers, the median provides the most meaningful measure of central tendency.

Control participant median completion time (12 minutes 47 seconds) was less than the intervention participant median completion time (17 minutes 31 seconds).

6.1.2.3 Fly or ground decision

Only in case 4 with the control participants was consensus in decision choice (*fly* or *ground*) seen. Figure 6.11 illustrates the differences of opinion, which were not obviously related to any explanatory factor.

A binomial test of proportions was carried out to test whether the intervention method resulted in a significantly different proportion of decisions toward *fly* using

	Alternate Choice / Last judgement use		Alternate Choice / experience (years)	
Case no.	R	p	R	p
1	-0.157	0.801	-0.758	0.138
2	0.721	0.17	0.292	0.634
3	0.784	0.116	0.931	0.022*
4	-0.721	0.17	-0.292	0.634
5	-0.419	0.482	-0.498	0.394
6	0.294	0.631	-0.152	0.808

Table 6.5: Correlations between intervention cohort decision (Ground at Gate 1, Ground Gate 2 and Fly) and participant factors.

the following hypotheses:

- H_0: there is no difference in the proportions of fly/ground decisions between control and intervention responses.
- H_a: there is a difference in the proportions of fly/ground decisions between control and intervention responses.

Comparing the proportion of *fly* decisions made by the control cohort ($n = \frac{13}{30}$), with the proportion of intervention cohort *fly* decisions ($n = \frac{17}{30}$), the difference was not significant at the 5% level: $N = 30, p = 0.143$. However, the small sample size has low statistical power, meaning that confidence in this conclusion should be improved by a larger experimental sample size. No significant correlations were found between the decision made and the experience or time since such a judgement was last made by the participants (Table 6.5).

6.1.2.4 Clarity of decision justifications

Some control participant decisions ($n = 9$) were difficult to decipher, although this view was subjective to the author. These particular responses made no clear statement of the decision choice, were populated with related but superfluous information to the immediate decision, or detailed the information necessary to

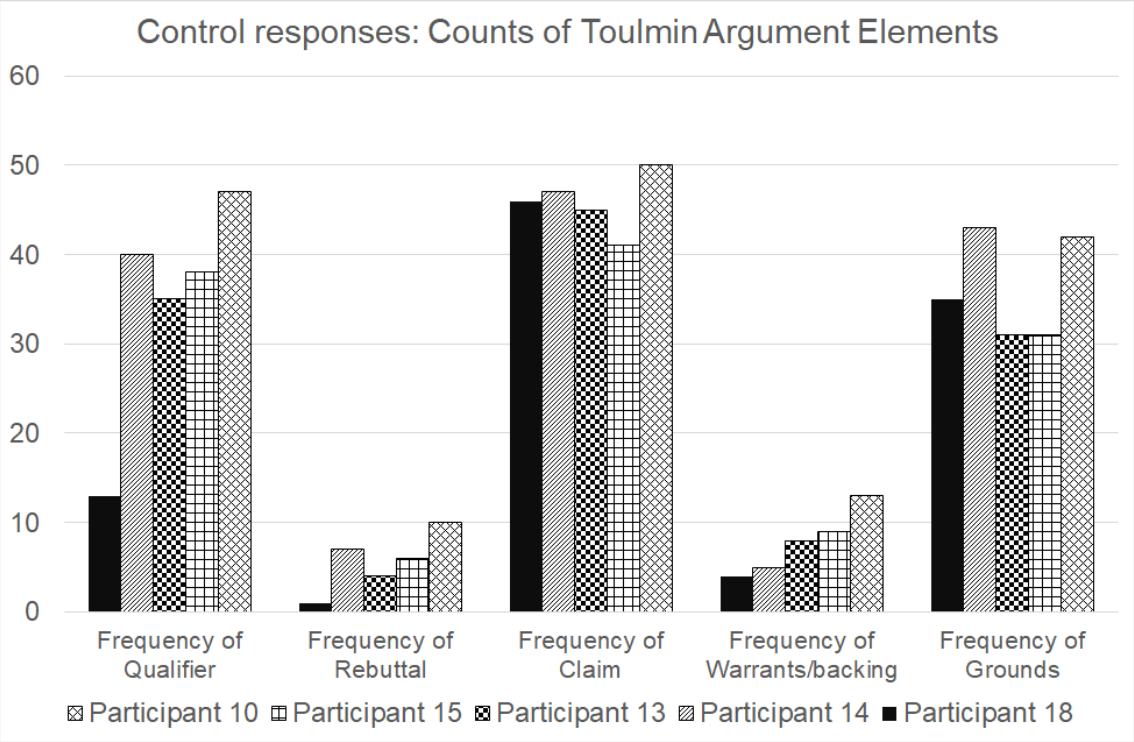


Figure 6.12: Counts of Toulmin schema elements within control participant decision justifications.

improve upon the decision basis without committing to a decision. Consequently, all control responses were analysed using the Toulmin schema as a coding structure, which enabled inferences to be made about the control group decision justification effectiveness. Although control participants were not instructed to form their arguments in a particular way, their decision justifications will have been performed to the best of their ability under the constraints of the experiment. On average, control responses mostly contained grounds, claims or qualifiers, but lacked warrants, backing and rebuttals, as shown in Figure 6.12.

Hypothesising that each claim requires a warrant (or backing) and rebuttal in order to substantiate the grounds of an argument, a count was made of the claims made within control responses. The proportion of claims made without a supporting warrant (or backing) was 82.6% (95% confidence interval is between 77.1% and 87.4%). The proportion of claims made without a rebuttal was 86.7% (95% confidence interval is between 81.5% and 90.8%). These results indicate

that the inductive arguments[6] made by participants are reliant upon suppressed premises to provide strength to their claim. Notably, everday arguments almost always depend upon suppressed premises (Sinnot-Armstrong and Fogelin, 2015, p96-101), but in a specific task to provide justification this result is surprising. It infers that an unstructured decision justification lacking in warrants/backing or conditions of rebuttal may not enable clear understanding of a reasoning behind practitioner judgements about structural integrity risk decisions, without knowledge of the decision circumstances.

6.2 Results from the second phase of the safety decision experiment

While the results from the first phase of the experiment, reported in Section 6.1, illustrate the usability and functionality of the proposed new framework, an independent assessment provides clearer evidence of the framework's potential advantages. An independent group of 11 experienced practitioners were recruited for the second phase of the experiment (referred to here as raters). They were tasked with comparing assessments made by the intervention cohort with assessments made by the control cohort. Three attributes were of interest: justifiability of the argument; transparency of the argument; and ease of understanding of the method used. Each rater considered six cases, meaning a total of 66 rater responses were collected for each of the three attributes of interest. While raters were not provided with training in the intervention method, they were given a brief context document (Appendix K), and broad details of each decision scenario only.

[6] An inductive argument is one where the premises are offered as a reason to accept the claim but not as a guarantee (Fulkerson, 1996, p13).

6.2.1 Primary results from the second phase

The primary aim of the comparison testing was to identify whether the novel framework provided a clearer audit trail of the judgement-based risk decision made about a structural integrity risk, compared to an unstructured control. This was achieved by paired comparison of control and intervention responses. Participants in the first phase of this experiment had been pre-assigned into *experience pairings*, which grouped control and intervention participants by experience levels to minimise confounding of the results.

The form of each comparison made was the direction (control or intervention preference) and magnitude of the preference[7]. For example, if the rater preferred the control participant's assessment *to a very great extent* for a particular attribute, the response variable would be -4. If the rater preferred the intervention participant's assessment *to some extent* for that attribute, the response variable would be +2.

With 11 raters making 6 comparisons each, there are $n = 66$ total paired comparisons for each of the three attributes. These data were analysed using Analysis of Variance (ANOVA). This method allowed the effect of the main variable of interest to be estimated (difference between intervention and control) while accounting for the effects of other experimental factors (the raters, the pairings and the cases).

The ANOVA model used[8] the rater identifier (*RaterID*), case number (*Case*) and experience pairing (*Pairing*) as "factors" in the model[9]. "Factors" in ANOVA account for the potential sources of experimental variability at the highest level of abstraction possible from the data. Considering the 11 participant raters in particular, the dependence within the 66 comparisons is accounted for by including *RaterID* as a factor in the model. The consequence of declaring the raters as a factor is a reduction in the degrees of the freedom available to the statistical analysis.

[7] Choices offered to raters were: To a very little extent / To some extent / To a great extent / To a very great extent.

[8] The ANOVA model used within Minitab (Minitab, 2020) was the *General Linear Model*.

[9] Captions in *(italicised brackets)* refer to the naming used within Tables 6.6, 6.7 and 6.8.

For each of the 66 samples, the response variable corresponds to the difference reported between the intervention and control members of each pair for each attribute (ease of understanding, greater transparency, and better justification). The output from the ANOVA test for each attribute is a general linear model that includes each factor as an independent variable. The constant coefficient of the general linear model is equal to the mean difference between the intervention and control responses. A positive constant indicates a mean preference towards the intervention, while a negative constant indicates mean preference towards the control. An overall mean value of zero would indicate no difference between the control and intervention, and the p-value refers to the test of the null hypothesis that the population mean difference is zero. This is the main result of interest since a value of the constant which is significantly different from zero provides evidence of a real difference between the novel framework and existing approaches.

In all three ANOVA models, the residual plots did not indicate any non-random patterns, which would have otherwise given cause for concern regarding the data meeting assumptions that validate the use of the model (Minitab, 2020).

6.2.1.1 Ease of Understanding

Participants were asked:

> "*Which, of the two decision justifications presented, is the easier to understand?*"
>
> Followed by:
>
> "*To what extent is the option you chose easier?* "

The ANOVA results are shown in Table 6.6. A p-value of less than 0.05 shows that *RaterID* has a significant association with the response variable *ease of understanding* at the 5% level. This provides evidence that the factor *RaterID* contributes to the variation in the response. The factors *Case* and *Pairing* are not significantly associated with *ease of understanding*. This is an ideal result since it illustrates that the experimental set-up did not influence the response variable - only the raters had an effect on the results.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
RaterID	10	133.931	13.3931	2.11	0.042
Case	5	37.677	7.5353	1.19	0.329
Pairing	4	2.923	0.7308	0.12	0.976
Error	46	291.319	6.3330		
Total	65	468.439			

Table 6.6: ANOVA table for the response variable “ease of understanding”, given the factors *RaterID*, *Case* and *Pairing*.

The constant coefficient in this model has a positive value (0.475) but the result is not significant at the 5% level $(p = 0.132)$. This result regarding the response variable *ease of understanding* is not generalisable, and the null hypothesis of no difference is not rejected.

6.2.1.2 Greater Transparency

ANOVA results for the *transparency* attribute are shown in Table 6.7. Participants were asked:

> “*Which, of the two decision justifications presented, provides the greater transparency of the decision basis provided by the author?*”
>
> Followed by:
>
> “*To what extent does the option you chose provide greater transparency?* ”

RaterID showed a very significant association at less than the 0.1% level, while the remaining factors, *Case* and *Pairing*, had no statistically significant association with the response variable at the 5% level.

The constant coefficient of the general linear model again indicated a positive value (0.883), but for this response variable the p-value was highly significant at the 0.1% level $(p = 0.001)$. This provides strong evidence to reject the null hypothesis and we can reasonably conclude that the novel framework is preferred over the existing approaches for showing *greater transparency*.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
RaterID	10	217.491	21.749	5.72	< 0.001
Case	5	19.270	3.854	1.01	0.421
Pairing	4	8.795	2.199	0.58	0.680
Error	46	174.963	3.804		
Total	65	417.030			

Table 6.7: ANOVA table for the response variable “greater transparency”, given the factors *RaterID*, *Case* and *Pairing*.

Source	DF	Adj SS	Adj MS	F-Value	P-Value
RaterID	10	138.800	13.880	2.80	0.009
Case	5	27.386	5.477	1.10	0.372
Pairing	4	5.758	1.440	0.29	0.883
Error	46	228.363	4.964		
Total	65	400.121			

Table 6.8: ANOVA table for the response variable “better justification”, given the factors *RaterID*, *Case* and *Pairing*.

6.2.1.3 Better Justification

Participants were asked:

> “*Which, of the two decision justifications presented, offers a better justification?*”
>
> Followed by:
>
> “*To what extent does the option you chose offer a better justification?*”

RaterID again had a significant association with the response variable at the 1.0% level, while *Case* and *Pairing* did not (Table 6.8).

The constant coefficient of the general linear model took the highest positive value of the three response variables at 1.240, which was highly significant at the 0.1% level ($p < 0.001$). This provides strong evidence to reject the null hypothesis and we can reasonably conclude that the novel framwork is preferred over existing approaches for providing a *better decision justification*.

	Intervention completeness	n	Median	Mean	σ
Ease of Understanding	Complete	45	2.00	0.82	2.56
	Partially Complete	11	2.00	0.64	2.84
	Incomplete	10	-2.50	-1.30	2.63
Greater Transparency	Complete	45	2.00	0.36	2.41
	Partially Complete	11	2.00	0.82	2.79
	Incomplete	10	0.00	-0.20	2.78
Better Justification	Complete	45	2.00	1.40	2.42
	Partially Complete	11	2.00	0.64	2.84
	Incomplete	10	2.50	1.20	2.53

Table 6.9: Descriptive statistics for the three response variables, arranged by the completeness level of the intervention assessment.

6.2.2 Secondary results from the second phase

Further analysis, including "intervention response completeness" as a factor in the ANOVA models did not indicate a significant association with any of the three attributes of interest. The means of the response variables for justification and transparency vary non-monotonically with the "intervention response completeness" (Table 6.9), but there is insufficient data in these sub-groups to form any general conclusions about the effect.

6.3 Novel framework testing discussion

The results in this chapter have provided evidence that the proposed new framework improves the communication of subjective judgements made to justify safety decisions concerning structural integrity faults and failures. First party feedback in the use of the proposed framework was wholly positive. Respondents felt guided through the application of their judgement, and the criticisms recorded may be the result of insufficient training for the experiment, and as a natural effect experienced by both intervention and control participants (explaining internalised processes). Third party comparisons favoured, on average, the

proposed framework, over the unstructured control, for its transparency and justification of arguments made in simulated naturalistic decision-making cases. These preferences indicate that the raters could understand the situation more clearly from the explanations given, and were more convinced by the intervention group's assessments compared to the control group's assessments. There is evidence therefore that the proposed framework enables a better audit trail to be established, for subjective judgement-based decisions about structural risks made under operational constraints. The experiment to test the new framework used engineering risk practitioners within the aviation domain and simulated naturalistic decision-making scenarios as far as possible. The experiment controlled variables as far as was possible to ensure that results would only reflect the judgements of participants and aid the delivery of well thought out conclusions (Fischhoff, 1992, p156).

6.3.1 Benefits of the proposed framework

This experiment has provided evidence that the proposed new framework benefits non-naive participants in constrained situations where they are unsupported by formal modelling or analysis and are compelled to make a deliberative subjective judgement-based decision. These are situations where a decision can not be deferred - operational situations are dynamic and a positive decision is necessary for continuity of performance or service delivery. Across six varied aviation maintenance test cases, the proposed framework has been flexibly applied by participants to justify their alternate choices ("fly" or "ground") about singular structural faults on one aircraft (cases 5 and 6) and in regards to wider systems considerations (cases 1 to 4). Although this demonstrates the flexibility of the method in a variety of engineering risk circumstances, this does not mean that the method may be generally applied as the wider system is likely to vary.

Time is known to influence decision-making in various ways (Klapproth, 2008), so future testing could vary the time limit for each scenario to test the limit of usability of the intervention method. As the time available decreases, the

opportunity for decision-maker deliberation reduces and the decision-making tends toward the intuitive decision-making already considered by the time-critical decision-making community (Okoli and Watt, 2018). There may also be an optimum time to permit unsupported judgement-based deliberation, up to which, decision-makers may have felt they required longer to justify their decisions (as seen in this experiment), and beyond which decision-makers may experience other cognitive effects. This requires further exploration.

Despite most participants expressing difficulties in being able to verbalise intrinsic assessments of a situation, all participants responded confidently and positively about the intervention method in operational situations. Time pressures and fluency issues with the framework were expected given the new method and relatively short training session:

- Intervention Participant SIRM-Phase4-2020_8 - "*The first one was a bit of a scramble, but I definitely felt I got better / faster at parsing the useful info from the non-useful info.* "

- Intervention Participant SIRM-Phase4-2020_6 - "*Having completed a 1-hour face-to-face (MS Teams) training session on the framework tool, I dived straight in to the first 3 exercises. I struggled a bit and have now taken a couple of hours to reflect before continuing. I think I have used the exercises so far as worked examples to understand how to use the tool. I now feel like more training would have helped.*"

- Intervention Participant SIRM-Phase4-2020_5 - "*The relationship between Grounds, Bridge and Support and how to apply the Voiding Conditions for certain arguments may not seem obvious at first and would require considerable training to standardise, particularly in junior and/or more inexperienced personnel.*"

These issues are expected to be overcome with improvements in training delivery and as familiarity with the method improves. Difficulties communicating internalised processing about the decision problems was a shared view with the

control participants, reinforcing the notion that it is difficult for humans to verbalise their intuition:

- Control participant SIRM-Phase4-2020_10, Case 1 - "*The remaining integrity of the mount due to an enlargement of the hole is speculated to be sufficient to retain the door, despite being OOL* [out of limits]. *I couldn't tell you why. Gut feeling. It would move a little more in the mount but hasn't been sufficient for it to be noticed in operation, or during latching.*"

- Intervention participant SIRM-Phase4-2020_6, Case 6 - "*I think I am taking credit in many of these exercises for damage tolerance in ac design which I cannot quantify. I am reproducing standard text extracts from the user guide but my reasoning is really about my experience of ac damage worse than these examples which does not deteriorate.*"

Few comments considered the heuristic decision tree specifically. Only participant SIRM-Phase4-2020_17 made reference about the decision process:

> "*The process of decision-making was logical, and followed instinctive thought patterns and progressions, considering key factors to the eventual decision.*"

Although the absence of evidence does not mean evidence of absence, the lack of specific comments suggests that the form of the heuristic decision tree did not create sufficiently visible issues to users. One would have expected negative experiences regarding the heuristic decision gates to have been mentioned. Targeted reflective questions may have extracted more actionable information.

The novel framework's compatibility with the situations and users it is intending to help will contribute to improving system resilience via organisational safety measures. The experiment tested judgement-based decision-making acceptability at the level of abstraction familiar to practitioners. Testing of the intervention protocol was limited to operational/end user applications, but as noted in the practitioner survey results, differences in risk analysis methodology were identifiable along respondent function and organisation category boundaries

(Chapter 4, Question 9d). Broader testing is therefore necessary to validate the utility of the intervention method in designer and regulatory environments. Similar constraints (time, information and resource) exist across these organisation categories (survey Question 25) but decision-making paradigms will differ between those environments. Validating usability in constrained situations attends to one of four models of resilience. If a method proves helpful to users when they are restricted to judgement-based decision-making, then it is *extending the system function boundaries* to adapt to an unforeseen condition, for which they otherwise would have remained unprepared (Dekker, 2019, p413).

This experiment confirmed the lack of quantified risk analysis applied by practitioners when under operational constraints identified from the results in Chapter 4. The novel framework intended to complement individual decision-makers in this regard by providing an approach that attended to subjective and situationally-specific criteria, rather than an approach that necessitated verbal or numerical quantification of the objective risk. The focus of the approach was on defining a sound inductive argument regarding the immediate *decision* about a hazard, using context-specific knowledge. Such knowledge is necessary for aviation maintenance decision-making (Mrusek and Douglas, 2020) and the intervention protocol aimed to support and structure deliberative reasoning, whether that be based upon qualitative or quantitative representations of risk. The results illustrate that this was able to be applied by practitioners under time constraints and with a means of structuring the justification.

An important observation is that the novel framework did not significantly change the proportion of *fly* decisions. This was an important finding of ethical significance. Advocating a method that encouraged more choices to “fly”, for example, might be reckless with obvious concerns for organisational safety and the future adoption of the novel framework. A broader sample base would add statistical power to this test. Changing the experiment to measure the novel framework influence on participants before and after intervention training would indicate individual participant effects.

6.3.1.1 Usability beyond the decision-maker

Within the context of risk management, effective justification aids "communication and consultation", "monitoring and review", and "recording and reporting", essential elements to the process advocated in the international risk management standard (The British Standards Institution, 2018a). Ensuring that risk activities are understandable will contribute to this goal, which as highlighted in Chapters 1 and 2, is currently lacking for unsupported judgement-based decision deliberations.

Given practitioners are compelled in some situations to make a decision either way, one objective was to support decision-makers by targeting improvement in the quality of risk decision justifications. The framework was designed with decision quality in mind (see Section 5.1.3) because other means of assuring decision quality such as expert consensus and normative optimisation decision methods (Wickens et al., 2013, p249) are not feasible under the constraints of naturalistic conditions.

Improvements in decision-making transparency would help third-parties in a number of ways: during safety incident investigation; when holding decision-makers to account; learning about structural integrity engineering decision-making for educating novices or improving policy. The decision-maker would also enhance their meta-cognitive awareness from a transparent feedback loop. Feedback from participants inferred some of these benefits:

– Intervention participant SIRM-Phase4-2020_8 - "*Finally it made me think with a structured approach, perhaps I am more risk averse than I expected.*"

– Intervention participant SIRM-Phase4-2020_6 - "*the process of mapping to the framework forced me to question the basis of each element, which was good.*"

The measure of information counts used is also indicative of improved auditability. The intervention responses not only used significantly more of the available case information, but also indicated a difference in the qualitative application of the information, through gross differences with control responses in the class pairings. While a greater quantity of information documented provides

a richer source from which to understand the decision-making circumstances post hoc, qualitative differences suggest information processing differences. For example, from the gross difference identified by the class pairing [I31, p1], this result indicates that intervention method participants were more likely to treat the situation analytically [decision class I31], given published limits for components at risk [case information class p1]. However, these were opportunistic results and the experiment was not designed to test differences in information classes. Further research would be necessary to investigate the extent of qualitative information usage differences between the intervention method and a control.

6.3.2 Control results

The control responses provide an interesting standalone data set. This experiment provided freedom for control participants to indicate how they would decide upon the presented test case situations. Despite the practitioner survey indicating that numerous risk assessment methods are used in operational situations to analyse structural integrity risks (Chapter 4, Question 9d), simulated scenario testing results have shown otherwise. No probabilities were used, no direct comparison with a probabilistic risk threshold was made, and no formalised analytical approach to risk analysis was applied in any of the 30 control decision responses. Participants were compelled to make an unsupported but deliberative judgement-based decision, and their output did not rely upon formal analysis. The responses are individually nuanced, with particular cues or shortfalls in available information and their influence on wider management considerations being highlighted:

> Control participant SIRM-Phase4-2020_14, case 6 - "*Composite structure may be subject to delamination and or moisture ingress and subsequent failure - especially pertinent on a control surface . Control surface replacement will also require rigging and system feedback checks to ensure the fg [flying] control system remains within control laws post the replacement.*"

This may be a consequence of the framing of the test case question, which

asked how the decision-maker would "handle this risk". Some responses were opaque in exactly what the decision outcome was, despite being informed in the context material of their role within the test cases:

> "*In all cases, you are an engineering manager responsible for aircraft structural integrity of the fleet. You are required to maintain the airworthiness of the fleet, providing balanced risk assessment of engineering faults against the operational requirement to fly, you are delegated with appropriate authorisations to make these decisions. You operate within your current regulatory and policy constraints. Your line management look to you to decide upon whether aircraft (individual or fleet wide) should be grounded, or limited in performance, in response to alerts about aircraft faults.*" (Appendix K)

Equally, unclear decision responses may be a consequence of the current role of some participants, who perhaps being in a more senior position than that they are being asked to role-play may operate on a higher level of abstraction than was required by the test cases. This was an unexpected effect of the experiment, despite all participants meeting the qualifying criteria.

6.3.3 Perception of acceptability

The test scenarios illustrated that decision-makers in both cohorts were, on average, very confident in presenting their judgement upon the decision. In complex risk contexts where access to meaningful analysis or feedback is not possible, such as those simulated in this experiment, decisions are unable to be objectively validated a priori. In these circumstances, the independent raters preferred, on average, the transparency and justification of responses made by intervention participants over the control.

It is apparent that decision-makers are not reliable judges of their own decision justification quality. There is already widespread evidence of human judgement fallibility (for example Fischhoff (1992, p147) or Bazerman and Moore (2009))

tending to undermine the trustworthiness of human perception for delivering rational outcomes. Despite having knowledge of normative "rational" analysis methods (Chapter 4, survey Question 9d), the experiment showed evidence of decision-makers that were satisficed in their approach - that their approach was good enough[10]. Advocates of bounded rationality as a suitable concept for decision-making highlight that under appropriate conditions, decision strategies that shortcut rigorous normative approaches may be suitable (Gigerenzer and Brighton (2009), Hafenbrädl et al. (2016)). Feedback from control participants made no mention of normative method (traditional rationality) dependence for decision-making, only a need for more information. Although, in the related work surrounding ecological rationality, this conclusion has been made under circumstances where information cue validity is known a priori (Gigerenzer, 2005). The individual perception of acceptability seen during the experiment (particularly among the control participants) highlights how a lack of meta-cognition may mislead individual decision-makers when deliberating structural integrity risks, especially where no analytical validation is available. This deductive finding supports the inferences made about the need to guide decision-makers making unsupported deliberative judgements (Chapter 4).

Intervention participants were not immune from the effect of overconfidence either, although intended to be mitigated by the use of argument to enable self-scrutiny of the decision-making basis, justifying premises and the claims. There were numerous cases of participants providing ambiguous support for premises, such as SIRM-Phase4-2020_16 for case 3 (Figure 6.13), who was very confident in their justification to *fly*.

Although the extract in Figure 6.13 was from a response categorised as a *complete* answer during analysis, the structure of this argument (and many others made by intervention participants) is logically inadequate. For example, the *support* does not provide evidence of why the *warrant* is acceptable and instead introduces a further *warrant* about the condition of the 3 brackets. This may have

[10]Chapter 4, Question 21 highlighted that practitioners use judgement when they perceived a solution to be "good enough".

Grounds: *the 3 remaining brackets provide sufficient support for the MFD should one fail.*		
Bridge: *load can be carried by the remaining 3 brackets*	*Support*: *There were no visible signs of damage to the 3 remaining brackets, confirmed by NDT*	
Voiding conditions: *failure is caused as a result of storage/handling/installation and more than one bracket has been damaged*		*Strength of support*: *Strong*
Claim: *Damage may not be common across ac, would be identified if it occured (low chance insidious failure) and there is sufficient load carried by remaining brackets in the event one fails*		

Figure 6.13: Participant SIRM-Phase4-2020_16 response to case 3.

been overlooked by the participant as a consequence of their inexperience with the tool, or misunderstanding of the function of the argument elements, both of which are reasons given in intervention participant feedback. However, this highlights the robustness of the intervention method through using inductive arguments, the raters seemed unaffected by logical inadequacies in their evaluative comparisons.

The intervention method's "Gate 3" was also intended to perform a "goal keeping" function, by challenging any inconsistencies, errors, or omissions within the decision-maker's argument basis. However, experimental limitations meant that the third gate could not be tested so participants were unable to be meaningfully challenged. The Gate 3 function was expected to help form a more complete argument, and as a feedback function it could improve risk-related skills (Fischhoff, 1992, p156) and decision-making attributes like sensemaking and situation awareness (Wickens et al., 2013, p247-249). This remains untested and further work to measure the benefits of the decision feedback using the intervention would enhance the assertion that this method improves safety resilience through the ability to "learn from the factual" (Dekker, 2019, p415).

6.4 Chapter summary

Simulation testing has confirmed the usability of a novel framework for practitioners deliberating aircraft maintenance structural integrity risk decisions. 5 pairs of participants made up the control and intervention testing cohorts across 6 written simulation scenarios under a time constraint of 20 minutes per scenario. Participants using the novel framework were very confident on average, but had difficulty in explaining intuitive reasoning about the scenarios and reported difficulties applying the argument categories to new and unfamiliar scenarios. Post hoc reflection identified some logical inconsistency of the justifications provided. The novel framework was compared with an unstructured control approach by 11 independent raters, who compared 6 pairs of decision justifications each and indicated their preferences on an ordinal Likert scale. ANOVA models provided statistical backing for the hypotheses that the novel framework provided a more

transparent and better justified decision in comparison to the control, but there was no difference for which approach was easier to understand. One key limitation of the experiment was that Gate 3 of the novel framework was unable to be tested.

The experiment also confirmed that under operational constraints of time, information and resource, practitioners do not attempt to apply normative decision-making methods to support deliberations about a risky prospect. Control decision responses demonstrated a variable quality of justification that tended to make use of suppressed premises and hidden background knowledge to support their decision choices. The novel method on the other hand directed users to document these elements, resulting in significantly higher information use. The novel decision support method did not significantly affect the proportion of fly or ground decisions selected.

Chapter 7

Conclusions

Maintenance managers make decisions almost every day on structural faults that occur throughout the service life of aircraft. They are required to balance availability demands for aircraft against the risk of continued flying given emergent structural damage, without guaranteed access to formal analysis. Situational constraints sometimes compel decision-makers to form unsupported judgement-based decisions about structural integrity faults and failures, which are deliberatively formed but not in accordance with recognised normative practices. No evidence of aircraft industry regulation has been found that provides sufficiently detailed guidance on the use of such subjective, deliberative judgements for risk decision-making.

The decision-making and risk analysis literature was reviewed for progress in this area. An abundance of work targeting risk assessment of structural damage and deterioration was found, but the methods used were intractable and incompatible with generalist users at the 'sharp end' of aircraft maintenance decision-making. There was a lack of consensus regarding the treatment of uncertainty in risk analysis, but a desire to improve collective scrutiny and decision-maker understanding of risk analysis by increasing transparency. These issues also related to challenges made to the credibility of risk analysis. Notwithstanding, simple risk management techniques, such as risk matrices, proliferate among practitioners, and are considered as a contributor to the lack of

transparency in the literature. However methods offered as solutions for improving the treatment of uncertainty and transparency, based on Bayesian updating, were not so prolific in practice owing to their requiring specialist knowledge to construct a requisite model. Consequently, a gap was noted in approaches for systematically handling risk that were usable by operators at the sharp-end where critical decision-making about risks are not just possible, but sometimes compelled under severe time pressure.

Argument and naturalistic decision-making were two specific areas that complemented the circumstances of the decision-making environment, the capabilities of the decision actors, and the desire to improve transparency in risk analysis. These were combined to develop a decision support framework for aircraft maintenance practitioners deliberating structural integrity risks. Improving support to decision-makers was noted to complement the growing safety science topic of resilience engineering, by increasing organisation robustness and by stretching system processes to handle surprises.

7.1 Satisfaction of aim and objectives

The aim of the study concerned aircraft maintenance decision-making about structural integrity risks:

– **Research project aim.** To support engineering decision-makers deliberating aircraft maintenance structural risks under constrained decision-making conditions.

This was broken into two supporting objectives:

– **Objective 1.** Gain a detailed understanding of the beliefs and practices of real-world practitioner decision-makers deliberating structural risks in operationally constrained conditions.

– **Objective 2.** Identify a method that supports decision-makers in forming a timely, effective, justifiable and robust judgement on unexpected aircraft structural damage in operational situations.

Survey responses from real-world engineering risk practitioners and three representative case studies informed the development of a novel framework to support decision-makers under operational constraints. The framework was tested using six representative aircraft maintenance scenarios, developed specifically for this purpose. Each one detailed a structural fault requiring a decision. Resources and information available to decision-makers was provided through a textual commentary, and a soft time limit of 20 minutes on the completion of each question simulated operational pressure.

The novel framework design was shaped by user requirements established from a survey of practicing engineering managers regarding structural integrity risk management, to ensure that issues reported were accounted for within the framework. Such issues included the difficulty of generating a meaningful risk likelihood, a lack of communication regarding risk appraisal aspects such as subjective judgement use, and differences across practitioner communities about the form that structural integrity risk should take.

7.1.1 Objective 1

The collection and analysis of primary and secondary data related to engineering maintenance structural integrity decision-making fulfilled the requirements of Objective 1. Primary data was collected via survey from experienced real-world practitioners, who were asked to respond to questions regarding structural integrity risk aspects when under operational constraints of time, information and resource. Additional data analysis supports the survey data, and was provided by three real-world case studies that regarded engineering maintenance incidents and decisions made under operational constraints. These results were reported and discussed in Chapters 3 and 4. The analysis showed that although risk concepts are valued in practical settings for managing structural integrity risk, practitioners experience difficulties in complying with risk management norms for structural faults.

One clear observation was that practitioners rely upon unsupported subjective

judgement when deliberating their decisions and actions in relation to structural faults, particularly when a situation compels a decision to be made under time pressure. In a minority of cases, the use of judgement and related attributes are not explicitly communicated during risk assessment and management. Insight into many different opinions regarding risk management across different organisations and roles was also afforded. Combining the reliance on subjective judgement with differing opinions highlights the opportunity for miscommunication and poor decision-making that may result from misaligned expectations about structural risk management activity.

The situational context also had critical bearing upon risk management activities and decision-making, evidenced, for example, by differences in risk analysis method choice across participant organisation types, and the difficulty some participants had in generating a likelihood. These observations were supported further by the novel framework testing experiment, which showed that control participants did not apply formal analytical risk approaches to support their decision choices.

This objective justifies the activity reported in Chapters 3 and 4, the findings from which helped inform the development and testing of the novel framework.

7.1.2 Objective 2

Using simulated operational maintenance scenarios to test safety decision-making, feedback from users of the novel framework provided confidence as to its usability under operational constraints. Comparison of the decision responses made using the framework against decision responses made using existing approaches provided statistical evidence that the framework improved justification and transparency of unsupported judgement-based decision rationales. These results address the fulfilment of Objective 2.

Given the finding that unsupported, subjective judgement-use is highly influential in decisions about structural integrity risk, the novel framework was designed with this as a central consideration. Since human judgement is known

to be fallible, the novel framework only permits decision actors to apply their unsupported judgement if there is a compelling reason to do so, mediated by the first node on the framework decision tree (Gate 1).

The form of the novel decision support framework was deliberately simple, using the principles of heuristic trees to establish a rapidly terminating decision flow, allowing timely attendance to faults while minimising the cognitive load on the decision-maker. The novel method also necessitated a structured argument to justify the decision in an unambiguous but flexible manner that was adaptable to the situation and the individual decision actor. The inductive form of argument advocated in the method means the premises do not guarantee the claim, but the claim is made more or less likely by the premises given. The resulting verbal reasoning in the argument is analogous to a probablistic risk interpretation but is entirely dependent upon the argument documented. Unlike traditional risk approaches, the novel framework does not require selection of numerical values to satisfy model requirements, but could optionally include quantitative information if the decision-maker considered its inclusion as contributing to the argument being made. In organisations that require quantitative risk approaches, this is helpful and a robust approach to assuring the justifiability of the risk within organisation settings that necessitate risk approaches. Although both control and intervention participants reported difficulties in verbalising intuitive perceptions of situations, the intervention model encouraged self-reflection of the decision bases and saw some participants justify their decision with clearer acknowledgement of their intuition.

Following testing, intervention participant decision-makers felt supported by the method, evidenced by intervention cohort responses that they felt “very confident” and that they “agree” the method would be usable in operational situations. Although control participants also responded similarly to the self-reflective perspectives despite their use of an unstructured approach, comparison of the decision justifications across the six test cases provided evidence that the intervention method responses were significantly (statistically) better in their transparency and quality of justification. For decision-makers considering structural integrity risks under operational constraints, Objective 2 is therefore considered to

have been met.

7.1.3 Satisfaction of the aim

Concluding that the supporting objectives have been met, it is natural to state that the aim has been met. However, agreement that the aim has been satisfied is conditional upon ontological perspectives about risk analysis. Some parties will perceive risk as an objective property that mathematically characterises uncertainty about a particular consequence occurrence and contains only limited (decision actor) value aspects. For these, the conclusion may be wholly unsatisfactory given the novel framework's subjective approach and the lack of mathematical treatment of the situation. However, a decision paradigm that frames the problem by the available choices from a decision actor's perspective adopts the complementary position. In this case, decision-maker value is key and risk is a social construction based upon the values eschewed by the decision-maker in that situation and time. Risk analysis is an input to the decision-making process, and given the evidence of practitioner inability to treat operationally constrained situations using risk analysis norms, addressing the problem from a decision paradigm perspective is entirely appropriate. Consequently, the evidence created by the research project indicates that the aim has been achieved.

7.2 Contributions to knowledge

The satisfaction of the research aim contributes to the body of knowledge in several ways:

- A survey of engineering managers has enabled a better understanding of real-world behaviours and industrial risk management practices regarding aircraft structural integrity under operational constraints.
- A novel framework has been developed, based upon heuristics and argumentation, to support decision-makers deliberating structural integrity risks under constraints of time, resource and information.

– Testing of the novel framework with experienced aircraft maintenance managers deliberating naturalistic decision-making scenarios, identified improved transparency and better justification when compared with existing unstructured risk decision approaches.

7.3 Implications and practical significance

For ‘sharp-end’ engineering practitioners with a compelling requirement to make a decision on a structural fault to some equipment that has entered service, the intervention provides a usable process and prompt with which to robustly justify and defend a judgement-based decision. Improved training is anticipated to overcome the errors of commission identified with intervention responses, but irrespectively, experimental testing evidenced improvements in the decision justification over an unstructured alternative.

Better justification of a decision output is of principal benefit to the decision-maker. Noting from the survey that practitioners receive influencing forces and pressure to make their decisions (survey Question 25), a decision justification constructed with an unambiguous argument provides robustness against criticism. In comparison, argument analysis of the unstructured control justifications showed that these responses were highly dependent on suppressed premises to be convincing. The use of argument in the novel method allows for improvement of the decision justification such that if a stakeholder challenges a particular element, a more complete argument will result from inclusion of the challenge and its supporting evidence.

A more *transparent* judgement-based decision justification has the potential to enhance decision efficacy through reflection. Firstly, for the decision-maker, a post-hoc review of the situation can offer a time-stamped defence of challenges that may be made with hindsight, once more information comes to light. Secondly, post hoc review may also exercise the feedback loop, allowing the decision-maker to review cues in the presence of subject experts or peers. The same process may also benefit novices looking to develop their mental models of real-world situations

from more experienced practitioners, or system designers who could use decision justifications to target improvements in the availability of maintenance information that operational decision-makers consider the most useful.

To the benefit of organisation managers, regulatory authorities and policy setters, the results draw attention to differences in the 'work as imagined' policies and 'work as done' practices for structural integrity risk decision-making. Survey responses evidence incompatibilities between decision-maker beliefs and organisational risk decision requirements, tension which may lead to miscommunication or misunderstandings about risky prospects. Policies encourage preferred behaviours among workers but where protocols are unworkable, as is sometimes the case with aircraft structural risk faults, clearer guidance is necessary to effectively support employee actions. The flexibility afforded by the novel framework could provide organisations with a procedurally rational approach for 'sharp-end' practitioners (on whom they depend, across a variety of situations, to use their judgement) when they are compelled to make decisions about structural risks. Although, as a note of caution, the method design is based on an assumption that professionals are not intent on manipulating situations through decision justifications for gain. For instance, one may easily apply the framework claiming the situation was as documented in their argument, while intentionally witholding contextual information that may have influenced the choices made.

The domain studied is safety centric, but the application of the principles applied need not be so restrictive and may find applications in domains where time-critical human (as opposed to system) decision performance is necessary. With further research, the method might be proven equally as supportive to decision-makers in situations with similar boundary conditions, namely: a decision required regarding a complex system; the decision alternatives can be simplified to a most basic binary and mutually exclusive form; constraints on resource, information and time prevent an objective assessment being completed; a human decision-maker is compelled to use their judgement and is accountable for their actions. Clearly, context-specific adjustment to the model wording would be necessary.

7.4 Limitations and future work

Notwithstanding restrictions imposed by government interventions throughout the course of the research project owing to the CoVID 19 pandemic and time pressures imposed by an additional requirements for ethical approval from the MODREC, there are improvement opportunities from this study.

Firstly, while the professional status of the author enabled privileged access to a body of practicing decision-makers, there are clear limitations regarding researcher axiology and detachment of personal values. Secondly, the lack of a second rater when carrying out thematic analyses increases the influence of subjectivity within the results. Thematic analyses influenced the development of the novel method and the results analysis of secondary effects about information use. The impact on the results analysis is therefore not considered substantial, but a second rater would enable more robust exploitation of the results discussed in Chapter 6. Thirdly, training of the intervention cohort during testing of the novel framework was conducted by the author as a consequence of the constraints of the academic award. Ideally, training would be conducted by an independent party to minimise the chances of participants being biased. Fourthly, Gate 3 of the heuristic decision tree was unable to be tested using the experimental method applied. The physical testing environment originally envisioned for testing, where another participant could then role play a reviewer at Gate 3 and challenge the original argument basis if deemed appropriate, was impeded by national pandemic restrictions. Disputes at Gate 3 could lead to tension regarding the final form of the argument, which could induce dynamic group influences on the decision-making, such as courtesy bias (Baybutt, 2018, p206). As these effects remain undocumented, Gate 3 interactions remain untested and therefore unverified, since online and non-interactive testing could only provide simulated and non-specific agreement at Gate 3. Lastly, and related to experimental set-up, the presentation of the simulations to practitioners was not representative of how information cues would be presented to a decision-maker in reality. Information flow to engineering managers would arrive as a mixture of verbal advice, written procedures, personal

knowledge, visual cues of faults, and operational influences such as urgency would likely be cultural in their manifestation. Given the circumstances the approach taken was a satisfactory compromise, but is nonetheless a limitation of the study.

Reflecting upon the CoVID 19 pandemic, the primary impact was to the experimental testing of the novel framework. UK government legal restrictions on social contact, local and national travel, and encouragement to work from home necessarily controlled public behaviours to minimise the spread of the virus but restricted the naturalistic experimental options open to the research project. The consequences were that a less engaging virtual platform of data collection was required to be used in Phase 4 for both prior training of participants and for all sub-phase data collection. This stretched the realism of the scenarios, where in-person delivery may have enabled physical damaged items to be handled by participants, as they might do in real-life.

Another restriction on the research project constituted the additional demands for ethical approval required by the MOD Research and Ethics Committee. Approval from the committee was required prior to the two data collection phases. Prior to the Phase 2 survey, unanticipated requirements to re-present to the committee for a second time caused a delay of 4 months before approval was given. The survey was originally intended to ground early perceptions of the research problem in a rigorous confirmation of aircraft maintenance practitioner beliefs and attitudes toward structural integrity risks. The anticipated delays meant that the survey scope was expanded to enable shaping of the novel framework development by including questions that explored the decision-making processes. Follow-up amendment to the initial ethical approval to include the Phase 4 experimental testing also required rework to satisfy the committee, although the delay at this stage was two weeks and had minor impact.

Increasing the quantity of participants would provide greater statistical confidence in the novel framework and inductive conclusions from the survey. The recruitment met the minimum necessary for meaningful statistical analysis, but this fell below personal expectations and highlighted the difficulty of attracting practicing professionals to a niche study. A greater reward incentive may have

improved uptake, but equally this may have increased unsuitable candidates. The critical reader may consider that low recruitment may be a consequence of overly restricted research scope. Author experience of the military aviation domain confidently rebuts this assertion in the knowledge that all similarly qualified engineering practitioners in the UK military have been required to make decisions typified by this study.

What has not been actively considered within this research program is the influence that safety culture and organisational limitations may have on the usability of the proposed framework. The success of a subjective decision approach depends on the risk behaviour of decision-makers within the organisation, which is shaped by the different cultural communities (Hillson, 2013). The constraints of the maintenance processes used in an organisation may also influence the novel framework's usability, where contractual limitations may inhibit working outside of written processes without approval. Further work exploring the influence of the method to differing organisation safety characteristics may benefit the usability of the framework.

The generalisability of the results are limited to aircraft maintenance decision-makers. The survey (Chapter 4) experienced difficulty in recruiting a sufficiently diverse participant base, and the novel method testing (Chapter 6) scope was required to be limited to a specific cohort to meet the objectives of testing in a specified environment. A majority of the participants in the survey identified as "aviation" practitioners, and the novel method testing was restricted to solely aviation maintenance decision-makers. However, as already discussed, the approach could be tailored for other domains. Future work in the area of naturalistic decision-making where actors are accountable may benefit from exploring deliberative decisions using the Toulmin argument schema.

7.5 Final thoughts

This project has exposed the vulnerabilities of structural integrity risk decision-making in aviation maintenance. Testing of the proposed framework

indicated usability and improved support for decision-makers communicating a judgement-based decision justification that is usable by practitioners in operationally constrained situations. The research objectives were set to provide greater support to decision-makers who are compelled to make such choices about structural integrity risks under operational constraints and without access to formal analytical solutions. The behaviours and beliefs of real-world practitioners collected via survey identified a lack of common approach regarding structural integrity risk treatment in operational situations. Differences of opinion about the efficacy and completeness of risk assessments made about structural faults were also apparent. Differences between organisation expectations about risk management highlighted where regulators, policy makers and managers lack the provision of appropriate guidance for decision-makers deliberating structural integrity risk.

The proposed framework is based upon heuristic decision gates and argument-based justification, which was designed to complement user capabilities and be flexible to the situation to provide procedural rationality. The subjective decision support tool enabled a boundedly rational actor to justify their decision about a risky situation in which a choice was compelled. The framework output is fundamentally different to a traditional risk measure, by being an argument about the soundness of a decision evaluation, rather than a quantity.

Decision-makers that participated in the experiment were confident using the novel framework and had belief in its usability in operational situations. Comparison of decision responses made using the novel framework with decision responses made using existing approaches gave evidence of improvements in transparency and in the quality of justification. Although the research project necessitated a compromised experimental method, which resulted in the proposed framework not being exposed to testing in its entirety, the major objectives of the research project have been achieved. Future work is encouraged in testing the method in other domains and industries that experience characteristically similar conditions: information, resource and time constraints; non-expert decision-makers; a compelling need for a decision; and a complex system being deliberated on. This

project considered structural integrity faults and failures in aircraft maintenance settings, but wider applicability is certainly possible and indeed hoped for in order to realise better support for decision-makers compelled to use their judgement in deliberating risky prospects.

References

Ale, B., Hartford, D., and Slater, D. (2015). ALARP and CBA all in the same game. *Safety Science*, 76:90–100.

American Psychological Association (2020a). Procedural Rationality. Available from: `https://dictionary.apa.org/procedural-rationality` [Accessed 26 May 2021].

American Psychological Association (2020b). Substantive Rationality. Available from: `https://dictionary.apa.org/substantive-rationality` [Accessed 26 May 2021].

Animah, I. and Shafiee, M. (2018). Condition assessment, remaining useful life prediction and life extension decision making for offshore oil and gas assets. *Journal of Loss Prevention in the Process Industries*, 53:17–28.

Apostolakis, G. E. (2004). How useful is quantitative risk assessment? *Risk Analysis*, 24(3):515–520.

Askeland, T., Flage, R., and Aven, T. (2017). Moving beyond probabilities - Strength of knowledge characterisations applied to security. *Reliability Engineering and System Safety*, 159(November 2016):196–205.

Aven, T. (2016a). On the use of conservatism in risk assessments. *Reliability Engineering and System Safety*, 146:33–38.

Aven, T. (2016b). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, 253(1):1–13.

Aven, T. (2016c). Supplementing quantitative risk assessments with a stage addressing the risk understanding of the decision maker. *Reliability Engineering & System Safety*, 152:51–57.

Aven, T. (2017). Improving risk characterisations in practical situations by highlighting knowledge aspects, with applications to risk matrices. *Reliability Engineering & System Safety*, 167(October 2016):42–48.

Aven, T. (2018a). An Emerging New Risk Analysis Science: Foundations and Implications. *Risk Analysis*, 38(5):876–888.

Aven, T. (2018b). How the integration of System 1-System 2 thinking and recent risk perspectives can improve risk assessment and management. *Reliability Engineering & System Safety*, 180(February 2017):237–244.

Aven, T. and Heide, B. (2009). Reliability and validity of risk analysis. *Reliability Engineering and System Safety*, 94(11):1862–1868.

Aven, T. and Zio, E. (2011). Some considerations on the treatment of uncertainties in risk assessment for practical decision making. *Reliability Engineering & System Safety*, 96(1):64–74.

Aven, T. and Zio, E. (2014). Foundational Issues in Risk Assessment and Risk Management. *Risk Analysis*, 34(7):1164–1172.

Baron, J. (2005). Normative Modesl of Judgment and Decision Making. In Koehler, D. J. and Harvey, N., editors, *Blackwell Handbook of Judgment and Decision Making*, chapter 2, pages 19–36. Blackwell, Oxford, 2nd edition.

Barons, M. J., Hanea, A. M., Wright, S. K., Baldock, K. C. R., Wilfert, L., Chandler, D., Datta, S., Fannon, J., Hartfield, C., Lucas, A., Ollerton, J., Potts, S. G., Carreck, N. L., Barons, M. J., Hanea, A. M., Wright, S. K., Baldock, K. C. R., Wilfert, L., Chandler, D., Datta, S., Fannon, J., Hartfield, C., Lucas, A., Potts, S. G., and Carreck, N. L. (2018). Assessment of the response of pollinator abundance to environmental pressures using structured expert elicitation. *Journal of Apicultural Research*, 57(5):593–604.

Baybutt, P. (2018). The validity of engineering judgment and expert opinion in hazard and risk analysis: The influence of cognitive biases. *Process Safety Progress*, 37(2):205–210.

Bazerman, M. H. and Moore, D. A. (2009). *Judgment in Managerial Decision Making*. John Wiley.

Beresford, B. and Sloper, T. (2008). *Understanding the Dynamics of Decision-Making and Choice : A Scoping Study of Key Psychological Theories to Inform The Design and Analysis of the Panel Study*. Social Policy Research Unit, University of York.

Berner, C. and Flage, R. (2016). Strengthening quantitative risk assessments by systematic treatment of uncertain assumptions. *Reliability Engineering and System Safety*, 151:46–59.

Bier, V. M. and Cox, L. A. (2007). Probabilistic Risk Analysis for Engineered Systems. In Edwards, W., Miles, R. F., and von Winterfeldt, D., editors, *Advances in Decision Analysis*, chapter 15, pages 279–301. Cambridge University Press, Cambridge, 1st edition.

Bjerga, T., Aven, T., and Zio, E. (2014). An illustration of the use of an approach for treating model uncertainties in risk assessment. *Reliability Engineering and System Safety*, 125:46–53.

Blockley, D. I. (1992). Setting the Scene. In Blockley, D. I., editor, *Engineering Safety*, chapter 1, pages 3–27. Cambridge University Press, Cambridge.

Bodde, M., van der Wel, K., Driessen, P., Wardekker, A., and Runhaar, H. (2018). Strategies for dealing with uncertainties in strategic environmental assessment: An analytical framework illustrated with case studies from The Netherlands. *Sustainability (Switzerland)*, 10(7):1–24.

Brown, C. L. and Utley, D. R. (2019). Ambiguity Aversion in Engineers. *Engineering Management Journal*, 31(1):2–7.

Brown, K. A. and Huyse, L. (2012). Why reliability-based approaches need more realistic corrosion growth modeling. In *Proceedings of the 2012 9th International Pipeline Conference*, pages 1–6, Calgary, Alberta.

Brown, R. V. (2015). Decision science as a by-product of decision-aiding: A practitioner's perspective. *Journal of Applied Research in Memory and Cognition*, 4(3):212–220.

Brugnach, M., Dewulf, A., Pahl-wostl, C., and Taillieu, T. (2010). Toward a Relational Concept of Uncertainty: about Knowing Too Little, Knowing Too Differently, and Accepting Not to Know [in press]. *Ecology and Society*, 13(2):1–17. Available from: https://www.ecologyandsociety.org/vol13/iss2/art30/ [Accessed 26 April 2021].

Buchak, L. (2010). Instrumental Rationality, Epistemic Rationality, and Evidence-Gathering. *Philosophical Perspectives*, 24(1):85–120.

Cartlidge, E. (2015). Why Italian earthquake scientists were exonerated. Science. Available from: http://www.sciencemag.org/news/2015/02/why-italian-earthquake-scientists-were-exonerated [Accessed 21 August 2018].

Chemweno, P., Pintelon, L., Muchiri, P. N., and Van Horenbeek, A. (2018). Risk assessment methodologies in maintenance decision making: A review of dependability modelling approaches. *Reliability Engineering and System Safety*, 173(June 2016):64–77.

Chen, L., Arzaghi, E., Abaei, M. M., Garaniya, V., and Abbassi, R. (2018). Condition monitoring of subsea pipelines considering stress observation and structural deterioration. *Journal of Loss Prevention in the Process Industries*, 51(December 2017):178–185.

Chionis, D. and Karanikas, N. (2018). Differences in Risk Perception Factors and Behaviours amongst and within Professionals and Trainees in the Aviation Engineering Domain. *Aerospace*, 5(2):62.

Civil Aviation Authority (2002). CAP 718: Human factors in aircraft maintenance and inspection. page 51. Available from: https://www.caa.co.uk/home/ [Accessed 26 April 2021].

Civil Aviation Authority (2003). CAP 716: Aviation Maintenance Human Factors (EASA / JAR145 Approved Organisations): Guidance Material on the UK CAA Interpretation of Part 145 Human Factors and Error Management Requirements. Available from: https://www.caa.co.uk/home/ [Accessed 26 April 2021].

Civil Aviation Authority (2015). CAP 795: Safety and Airspace Regulation Group Safety Management Systems (SMS) guidance for organisations. Available from: https://www.caa.co.uk/home/ [Accessed 26 April 2021].

Clemen, R. T. and Reilly, T. (2001). *Making Hard Decisions*. Duxbury, Pacific Grove, CA, 2nd edition.

Cohen, M. S., Freeman, J. T., and Wolf, S. (1996). Metarecognition in Time-Stressed Decision Making: Recognizing, Critiquing, and Correcting. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 38(2):206–219.

Conlin, H., Brabazon, P., and Lee, K. (2004). Exploring the Role and Content of the Safety Case. *Process Safety and Environmental Protection*, 82(4):283–290.

Cooke, R. M. (1991). *Experts in Uncertainty - Opinion and Subjective Probability in Science*. Oxford University Press, New York, 1st edition.

Cope, D. and Moffett, J. (2012). Bayesian Updating of Structural Health and Integrity Assessments using Real-Time Inspection Results. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, 20th AIAA/ASME/AHS Adaptive Structures Conference,14th AIAA*, pages 1–19, Reston, Virigina. American Institute of Aeronautics and Astronautics.

Coronavirus Act 2020 (2020). Coronavirus Act 2020, c. 7. Available from: https://www.legislation.gov.uk/ukpga/2020/7/introduction/enacted [Accessed 26 April 2021].

Cox, L. A. (2008). What's Wrong with Risk Matrices? *Risk Analysis*, 28(2):497–512.

Cox, L. A. (2012). Confronting Deep Uncertainties in Risk Analysis. *Risk Analysis*, 32(10):1607–1629.

Croskerry, P., Singhal, G., and Mamede, S. (2013a). Cognitive debiasing 1: Origins of bias and theory of debiasing. *BMJ Quality and Safety*, 22(Suppl. 2):58–64.

Croskerry, P., Singhal, G., and Mamede, S. (2013b). Cognitive debiasing 2: impediments to and strategies for change. *BMJ Quality & Safety*, 22(Suppl. 2):65–72.

Crossley, W. A., Luan, S., Allison, J. T., and Thurston, D. L. (2017). Optimization problem formulation framework with application to engineering systems. *Systems Engineering*, 20(6):512–528.

Dallat, C., Salmon, P. M., and Goode, N. (2019). Risky systems versus risky people: To what extent do risk assessment methods consider the systems approach to accident causation? a review of the literature. *Safety Science*, 119:266–279.

Day, A. (2020). Projects and Gamification Make Learning Fun. In *Decision Analysis Affinity Group Conference, 24*, page 11, Virtual. [Powerpoint presentation, available from: `https://www.decisionprofessionals.com/library/DAAG-Talks` accessed 26 April 2021].

Defence Safety Authority (2018). *DSA/SI/16/028: Service Inquiry - Griffin HT Mk1 ZJ241, Snowdonia - 09 August 2016*. UK Ministry of Defence, London. Available from: `https://www.gov.uk/government/publications/service-inquiry-into-the-accident-involving-a-griffin-mk1-zj241-at-yr-aran-snowdonia-wales` [Accessed 21 January 2019].

Dekker, S. (2019). *Foundations of Safety Science: A Century of Understanding Accidents and Disasters*. CRC Press, Boca Raton.

Dilger, R., Hickethier, H., and Greenhalgh, M. D. (2009). Eurofighter a safe life aircraft in the age of damage tolerance. *International Journal of Fatigue*, 31(6):1017–1023.

Durodié, B. (2017). Theory informed by practice. Application informed by purpose. Why to understand and manage risk, cultural context is the key. *Safety Science*, 99:244–254.

Eleye-Datubo, A. G., Wall, A., Saajedi, A., and Wang, J. (2006). Enabling a powerful marine and offshore decision-support solution through bayesian network technique. *Risk Analysis*, 26(3):695–721.

Elms, D. (1992). Risk Assessment. In Blockley, D. I., editor, *Engineering Safety*, chapter 2, pages 28–46. McGraw-Hill, Cambridge.

European Aviation Safety Agency (2018). Easy Access Rules for Airworthiness and Environmental Certification (Regulation (EU) No 748/2012).

Evans, J. S. B. T. (2019). Reflections on reflection: the nature and function of type 2 processes in dual-process theories of reasoning. *Thinking & Reasoning*, 25(4):383–415.

Faisandier, A., Roedler, G., Turner, R., Adcock, R., and Sofer, A. (2020). System requirements in SEBoK Editorial Board. In Cloutier, R., editor, *The Guide to the Systems Engineering Body of Knowledge (SEBoK). v.2.3*. Hoboken, NJ: The

Trustees of the Stevens Institute of Technology. https://www.sebokwiki.org/w/index.php?title=System_Requirements&oldid=59924 [Accessed 5 April 2021]. BKCASE is managed and maintained by the Stevens Institute of Technology Systems Engineering Research Center, the International Council on Systems Engineering, and the Institute of Electrical and Electronics Engineers Systems Council.

Fawaz, S. A. (2003). Equivalent initial flaw size testing and analysis of transport aircraft skin splices. *Fatigue and Fracture of Engineering Materials and Structures*, 26(3):279–290.

Federal Aviation Authority (2011). *AC 120-104: Establishing and Implementing Limit of Validity To Prevent Widespread Fatigue Damage*. US Department for Transport, Washington DC, USA. Available from: https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/638510 [Accessed 26 April 2021].

Fischhoff, B. (1992). Risk taking: a developmental prospective. In Yates, J. F., editor, *Risk-taking Behavior*, chapter 5, pages 133–162. John Wiley & Sons, Inc., Chichester, West Sussex.

Foddy, W. (1993). *Constructing Questions for Interviews and Questionnaires - Theory and Practice in Social Research*. Cambridge University Press, Cambridge, 1st edition.

Frangopol, D. M. (2011). Life-cycle performance, management, and optimisation of structural systems under uncertainty: accomplishments and challenges 1. *Structure and Infrastructure Engineering*, 7(6):389–413.

Fulkerson, R. (1996). *Teaching the Argument in Writing*. National Council of Teachers of English, USA.

Gainsburg, J., Fox, J., and Solan, L. M. (2016). Argumentation and Decision Making in Professional Practice. *Theory into Practice*, 55(4):332–341.

Gernand, J. M. (2018). A Set of Preliminary Model Experiments for Studying Engineering Student Biases in the Assessment and Prioritization of Risks. volume 13: Design, Reliability, Safety, and Risk of *ASME International Mechanical Engineering Congress and Exposition*. V013T05A021.

Gigerenzer, G. (2005). Fast and Frugal Heuristics. In Koehler, D. J. and Harvey, N., editors, *Blackwell Book of Judgement and Decision Making*, chapter 4, pages 62–88. Blackwell, Padstow, UK.

Gigerenzer, G. (2014). *Risk Savvy: How to Make Good Decisions*. Penguin Books, London, kindle edition.

Gigerenzer, G. and Brighton, H. (2009). Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, 1(1):107–143.

Gigerenzer, G., Todd, P. M., and Group, T. A. R. (1999). *Simple Heuristics That Make Us Smart*. Oxford University Press, Oxford.

Goerlandt, F., Khakzad, N., and Reniers, G. (2017). Validity and validation of safety-related quantitative risk analysis: A review. *Safety Science*, 99:127–139.

Gollwitzer, S. and Rackwitz, R. (1990). On the reliability of Daniels systems. *Structural Safety*, 7(2-4):229–243.

Gore, J., Flin, R., Stanton, N., and Wong, B. L. (2015). Applications for naturalistic decision-making. *Journal of Occupational and Organizational Psychology*, 88(2):223–230.

Govier, T. (2010). *A Practical Study of Argument*. Wadsworth, Cengage Learning, Belmont, USA, 7th edition.

Grant, E., Salmon, P. M., Stevens, N. J., Goode, N., and Read, G. J. (2018). Back to the future: What do accident causation models tell us about accident prediction? *Safety Science*, 104(December 2017):99–109.

Green, R., McNaught, K., and Saddington, A. (2021). Aircraft maintenance decision-making with unsupported judgement under operational constraints. *Unpublished manuscript*, pages 1–35.

Groarke, L. (2020). Informal Logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2020 edition.

Guthrie, R. and Reid, E. (2018). Estimating Landslide Induced Probability of Failure to Pipelines Using a Structured Reductionist Approach. In *Volume 2: Pipeline Safety Management Systems; Project Management, Design, Construction, and Environmental Issues; Strain Based Design; Risk and Reliability; Northern Offshore and Production Pipelines*, pages 1–8. ASME.

Haddon-Cave, C. (2009). *The Nimrod Review: An independent review into the broader issues surrounding the loss of rhe RAF Nimrod MR2 Aircraft XV230 In Afghanistan in 2006*. The Stationary Office, London,UK. Available from: `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/229037/1025.pdf`.

Hadorn, G. H. (2016). Temporal Strategies for Decision-making. In Hansson, S. O. and Hirsch Hadorn, G., editors, *The Argumentative Turn in Policy Analysis*, volume 10 of *Logic, Argumentation & Reasoning*, pages 217–242. Springer International Publishing, Cham.

Hafenbrädl, S., Waeger, D., Marewski, J. N., and Gigerenzer, G. (2016). Applied Decision Making With Fast-and-Frugal Heuristics. *Journal of Applied Research in Memory and Cognition*, 5(2):215–231.

Hagmann, J. (2012). Fukushima: Probing the analytical and epistemological limits of risk analysis. *Journal of Risk Research*, 15(7):801–815.

Hansson, S. O. and Aven, T. (2014). Is Risk Analysis Scientific? *Risk Analysis*, 34(7):1173–1183.

Hansson, S. O. and Hirsch Hadorn, G. (2018). Argument-based decision support for risk analysis. *Journal of Risk Research*, 21(12):1449–1464.

Hardman, D. K. and Ayton, P. (1997). Arguments for Qualitative Risk Assessment: The StAR Risk Adviser. *Expert Systems*, 14(1):24–36.

He, R., Li, X., Chen, G., Wang, Y., Jiang, S., and Zhi, C. (2018). A quantitative risk analysis model considering uncertain information. *Process Safety and Environmental Protection*, 118:361–370.

Hiles, A. (2012). *The Definitive Handbook of Business Continuity Management*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 3rd edition.

Hillson, D. (2013). The A-B-C of risk culture: how to be risk-mature. In *PMI® Global Congress*, New Orleans. Project Management Institute.

Hillson, D. (2017). *Understanding and Managing Risk Attitude*. Routledge.

Hoffman, P. C. (2009). Fleet management issues and technology needs. *International Journal of Fatigue*, 31(11-12):1631–1637.

Hoffrage, U. and Marewski, J. N. (2015). Unveiling the Lady in Black: Modeling and aiding intuition. *Journal of Applied Research in Memory and Cognition*, 4(3):145–163.

Howard, R. A. and Abbas, A. E. (2016). *Foundations of Decision Analysis (Global Edition)*. Pearson Education, Harlow, 1st edition.

Hurtado, J. L. and Hoffman, P. (2006). Airframe Integrity Based on Bayesian Approach. In *Proceedings - Annual Reliability and Maintainability Symposium*, pages 630–635.

Hurteau, M., Rahmanian, J., Houle, S., and Marchand, M. P. (2020). The Role of Intuition in Evaluative Judgment and Decision. *American Journal of Evaluation*, 41(3):326–338.

Isaksen, B. G. M. and McNaught, K. R. (2019). Uncertainty handling in estimative intelligence - challenges and requirements from both analyst and consumer perspectives. *Journal of Risk Research*, 22(5):643–657.

JASP Team (2019). JASP (Version 0.11.1) [Computer software].

Jordaan, I. (2005). *Decisions under Uncertainty: Probabilistic Analysis for Engineering Decisions*. Cambridge University Press, Cambridge.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291.

Khakzad, N., Khan, F., and Amyotte, P. (2011). Safety analysis in process facilities: Comparison of fault tree and Bayesian network approaches. *Reliability Engineering & System Safety*, 96(8):925–932.

Khan, F., Rathnayaka, S., and Ahmed, S. (2015). Methods and models in process safety and risk management: Past, present and future. *Process Safety and Environmental Protection*, 98:116–147.

Kiureghian, A. and Ditlevsen, O. (2009). Aleatoric or Epistemic? Does it matter? *Struct Safety*, 31(2):105–112.

Klapproth, F. (2008). Time and decision making in humans. *Cognitive, Affective and Behavioral Neuroscience*, 8(4):509–524.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, 50(3):456–460.

Kolmogorov, A. (2018). *Foundations of the Theory of Probability*. Dover Publications, New York, 2nd eng edition.

Krause, P., Fox, J., and Judson, P. (1995). Is There a Role for Qualitative Risk Assessment? *ArXiv*, abs/1302.4970.

Lambert, B. (2018). *A Student's Guide to Bayesian Statistics*. SAGE Publications Ltd, London.

Langseth, H. and Portinale, L. (2007). Bayesian networks in reliability. *Reliability Engineering & System Safety*, 92(1):92–108.

Lathrop, J. and Ezell, B. (2017). A systems approach to risk analysis validation for risk management. *Safety Science*, 99:187–195.

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science*, 42(4):237–270.

Li, C., Mahadevan, S., Ling, Y., Choze, S., and Wang, L. (2017). Dynamic Bayesian Network for Aircraft Wing Health. *AIAA Journal*, 55(3):930–941.

Lincoln, Y. S. and Guba, E. G. (1985). *Naturalistic Inquiry*. SAGE Publications Ltd, London.

Lindley, D. V. (2014). *Understanding Uncertainty*. John Wiley & Sons, Inc., Hoboken, NJ, USA, revised edition.

Ludlow, T. (2020). A decision archetype framework that helps individuals to make significant decisions quickly. In *Decision Analysis Affinity Group Conference, 24-26 June 2020*. Society of Decision Professionals. Available from: `https://www.decisionprofessionals.com/` [Accessed 26 May 2021].

Ma, Y., Wang, L., Zhang, J., Xiang, Y., and Liu, Y. (2014). Bridge Remaining Strength Prediction Integrated with Bayesian Network and In Situ Load Testing [in press]. *Journal of Bridge Engineering*, 19(10):1–11.

Maldonato, M., Dell'Orco, S., and Sperandeo, R. (2017). When intuitive decisions making, based on expertise, may deliver better results than a rational, deliberate approach. *Smart Innovation, Systems and Technologies*, 69:369–377.

Marewski, J. N., Gaissmaier, W., and Gigerenzer, G. (2010). Good judgments do not require complex cognition. *Cognitive processing*, 11(2):103–121.

Marewski, J. N. and Gigerenzer, G. (2012). Heuristic decision making in medicine. *Dialogues in clinical neuroscience*, 14(1):77–89.

Mellors, B. and Locke, C. (2007). What Have We Learned from Our Mistakes. In Edwards, W., Miles, R. F., and von Winterfeldt, D., editors, *Advances in Decision Analysis*, chapter 18, pages 351–374. Cambridge University Press, Cambridge, 1st edition.

Merkhofer, M. W. (1987). Quantifying judgmental uncertainty: Methodology, experiences, and insights. *IEEE Transactions on Systems, Man, and Cybernetics*, 17(5):741–752.

Microsoft Corporation (2020). Microsoft Excel for Office 365 MSO (version 16.0.12527.21230) 64-bit.

Military Aviation Authority (2015). *Regulatory Article 1210 (Issue 3) - Ownership and Management of Operating Risk (Risk to Life)*. UK Ministry of Defence, London, UK. Available from: `https://www.gov.uk/government/collections/maa-regulatory-publications` [Accessed 26 April 2021].

Military Aviation Authority (2016). *Regulatory Article 4812 (Issue 4) - Certification of Aircraft and Component Release (MRP 145.A.50)*. UK Ministry of Defence, London, UK. Available from: `https://www.gov.uk/government/collections/maa-regulatory-publications` [Accessed 26 April 2021].

Military Aviation Authority (2018a). *Regulatory Article 1210 (Issue 5): Ownership and Management of Operating Risk (Risk to Life)*. UK Ministry of Defence, London, UK. Available from: `https://www.gov.uk/government/publications/regulatory-article-ra-1210-ownership-and-management-of-operating-risk-risk-to-life` [Accessed 22 January 2019].

Military Aviation Authority (2018b). *Regulatory Article 1230 (Issue 5): Design Safety Targets Compliance Compliance*. UK Ministry of Defence, London, UK. Available from: `https://www.gov.uk/government/publications/regulatory-article-ra-1230-design-safety-targets` [Accessed 22 January 2019].

Military Aviation Authority (2019). *Manual of Air System Safety Cases (MASSC)*. UK Ministry of Defence, London, UK. Available from: https://www.gov.uk/government/publications/manual-of-air-system-safety-cases-massc [Accessed 26 April 2021].

Military Aviation Authority (2020a). *MAA02: Military aviation authority (MAA) Master Glossary*. UK Ministry of Defence, London, UK. Available from: https://www.gov.uk/government/publications/maa02-military-aviation-authority-maa-master-glossary [Accessed 26 April 2021].

Military Aviation Authority (2020b). *Regulatory Article 4812 (Issue 6) - Certification of Air System Release and Component Release*. UK Ministry of Defence, London, UK. Available from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/939286/RA4812_Issue_6.pdf [Accessed 26 April 2021].

Military Aviation Authority (2020c). *Regulatory Article 4813 (Issue 6): Maintenance Records (MRP 145.A.55)*. UK Ministry of Defence, London, UK. Available from: https://www.gov.uk/government/publications/regulatory-article-ra-4813-maintenance-records-mrp-145a55 [Accessed 26 April 2021].

Military Aviation Authority (2020d). *Relevant to Regulatory Article 5726 (Issue 2): integrity management*. UK Ministry of Defence, London, UK. https://www.gov.uk/government/publications/regulatory-article-ra-5726-integrity-management [Accessed 26 April 2021].

Minitab, L. (2020). Interpret the key results for Fit General Linear Model. https://support.minitab.com/en-us/minitab/19/help-and-how-to/statistical-modeling/anova/how-to/fit-general-linear-model/interpret-the-results/key-results/?SID=113267 [Accessed 26 April 2021].

Montibeller, G. (2020). Fighting against cognitive, motivational and group biases in decision analysis. Presentation available from https://www.decisionprofessionals.com/articles/Fighting-against-cognitive-motivational-and-group-biases-in-Decision-Analysis [Accessed 26 April 2021].

Montibeller, G. and von Winterfeldt, D. (2015). Cognitive and Motivational Biases in Decision and Risk Analysis. *Risk Analysis*, 35(7):1230–1251.

Morgan, J. I., Abbott, R., Furness, P., and Ramsay, J. (2016). UK rail workers' perceptions of accident risk factors: An exploratory study. *International Journal of Industrial Ergonomics*, 55:103–113.

Mosier, K. L. (2008). Technology and "Naturalistic" Decision Making: Myths and Realities. In Schraagen, J. M., Militello, L. G., Ormerod, T., and Lipshitz, R.,

editors, *Naturalistic Decision Making and Macrocognition*, chapter 3, pages 41–54. 1st edition.

Mousavi, S. and Gigerenzer, G. (2014). Risk, uncertainty, and heuristics. *Journal of Business Research*, 67(8):1671–1678.

Mrusek, B. and Douglas, S. (2020). From Classroom to Industry : Human Factors in Aviation Maintenance. *University Aviation Association*, 38(2).

Murphy, D. M. and Pate-Cornell, M. E. (1996). The SAM Framework: Modeling the Effects of Management Factors on Human Behavior in Risk Analysis. *Risk Analysis*, 16(4):501–515.

Naweed, A. and Kingshott, K. (2019). Flying Off the Handle: Affective Influences on Decision Making and Action Tendencies in Real-World Aircraft Maintenance Engineering Scenarios. *Journal of Cognitive Engineering and Decision Making*, 13(2):81–101.

Neumann, P. J. and Politser, P. E. (1992). Risk and optimality. In Yates, F. J., editor, *Risk-taking Behavior*, chapter 2, pages 27–47. John Wiley & Sons, Inc., Chichester, West Sussex, 1st edition.

Nielsen, J. J. and Sørensen, J. D. (2010). Bayesian networks as a decision tool for O&M of offshore wind turbines. *ASRANet: Integrating Structural Analysis, Risk & Reliability*, (1):1–8.

Oberkampf, W. and Helton, J. (2002). Investigation of evidence theory for engineering applications. In *43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, pages 1–15.

Okoh, C., Roy, R., Mehnen, J., and Redding, L. (2014). Overview of Remaining Useful Life prediction techniques in Through-life Engineering Services. *Procedia CIRP*, 16:158–163.

Okoli, J. and Watt, J. (2018). Crisis decision-making: the overlap between intuitive and analytical strategies. *Management Decision*, 56(5):1122–1134.

Osmo, R. and Benbenishty, R. (2004). Children at risk: Rationales for risk assessments and interventions. *Children and Youth Services Review*, 26(12):1155–1173.

Over, D. (2005). Rationality and the Normative/Descriptive Distinction. In Koehler, D. J. and Harvey, N., editors, *Blackwell Handbook of Judgment and Decision Making*, chapter 1, pages 4–18. Blackwell, Oxford, 2nd edition.

Pasman, H. and Rogers, W. (2018). How trustworthy are risk assessment results, and what can be done about the uncertainties they are plagued with? *Journal of Loss Prevention in the Process Industries*, 55(May):162–177.

Pasman, H. J., Rogers, W. J., and Mannan, M. S. (2017). Risk assessment: What is it worth? Shall we just do away with it, or can it do a better job? *Safety Science*, 99:140–155.

Patankar, M. S. and Taylor, J. C. (2004). *Risk Management and Error Reduction in Aviation Maintenance*. Ashgate Publishing Ltd, Abingdon.

Pate-Cornell, M. E. (2007). The Engineering Risk Analysis Method and Some Applications. In Edwards, W., Miles, R. F., and von Winterfeldt, D., editors, *Advances in Decision Analysis*, chapter 16, pages 302–324. Cambridge University Press, Cambridge, 1st edition.

Payne, J. W., Bettman, J. R., and Johnson, E. J. (1993). *The Adaptive Decision Maker*. Cambridge University Press, Cambridge.

Peterson, M. (2009). *An Introduction to Decision Theory*. Cambridge University Press, Cambridge, 1st edition.

Phillips, J. K., Klein, G., and Sieck, W. R. (2005). Expertise in Judgment and Decision Making: A Case for Training Intuitive Decision Skills. In Koehler, D. J. and Harvey, N., editors, *Blackwell Handbook of Judgment and Decision Making*, chapter 15, pages 297–315. Blackwell, Padstow, UK, 2nd edition.

Pidgeon, N. (1998). Risk assessment, risk values and the social science programme: why we do need risk perception research. *Reliability Engineering & System Safety*, 59(1):5–15.

Pitz, G. F. (1992). Risk taking, design, and training. In Yates, J. F., editor, *Risk Taking Behaviour*, chapter 10, pages 283–320. John Wiley & Sons, Inc., Chichester, West Sussex.

Polacsek, T., Sharma, S., Cuiller, C., and Tuloup, V. (2018). The need of diagrams based on Toulmin schema application: an aeronautical case study. *EURO Journal on Decision Processes*, 6(3-4):257–282.

Proenca, D., Estevens, J., Vieira, R., and Borbinha, J. (2017). Risk Management: A Maturity Model Based on ISO 31000. In *2017 IEEE 19th Conference on Business Informatics (CBI)*, volume 1, pages 99–108. IEEE.

Provan, D. J., Rae, A. J., and Dekker, S. W. (2019). An ethnography of the safety professional's dilemma: Safety work or the safety of work? *Safety Science*, 117(April):276–289.

Public Records Act 1958 (1958). Public Records Act 1958. Available from: `http://www.legislation.gov.uk/ukpga/Eliz2/6-7/51` [Accessed 29 June 2020].

Qualtrics (2020). Qualtrics. Available from: `https://www.qualtrics.com/` [Accessed 26 May 2021].

Rae, A. and Alexander, R. (2017). Forecasts or fortune-telling: When are expert judgements of safety risk valid? *Safety Science*, 99:156–165.

Rae, A., Alexander, R., and McDermid, J. (2012). The Science and Superstition of Quantitative Risk assessment. In *Proceedings of Probabilistic Safety Assessment and Management (PSAM) Conference 11, International Association for Probabilistic Safety Assessment and Management (IAPSAM), Helsinki, June 2012*, pages 2292–2301.

Rae, A., Alexander, R., and McDermid, J. (2014). Fixing the cracks in the crystal ball: A maturity model for quantitative risk assessment. *Reliability Engineering & System Safety*, 125:67–81.

Rae, A., Provan, D., Aboelssaad, H., and Alexander, R. (2020). A manifesto for Reality-based Safety Science. *Safety Science*, 126(September 2019):1–11.

Rafiq, M. I., Chryssanthopoulos, M. K., and Sathananthan, S. (2015). Bridge condition modelling and prediction using dynamic Bayesian belief networks. *Structure and Infrastructure Engineering*, 11(1):38–50.

Rail Accident Investigation Branch (2016). *Rail Accident Report 22/2016: Structural failure caused by scour at Lamington viaduct, South Lanarkshire 31 December 2015*. Rail Accident Investigation Branch, Derby, UK. Available from: `https://www.gov.uk/raib-reports/structural-failure-at-lamington-viaduct` [26 April 2021].

Ramiah, S. P. and Banks, A. P. (2015). Naturalistic decision making through argumentation: Resolving labour disputes. *Journal of Occupational and Organizational Psychology*, 88(2):364–381.

Ramsey, F. P. (1926). The Foundations of Mathematics. *Proceedings of the London Mathematical Society*, s2-25(1):338–384.

Rasmussen, J. (1997). Risk Management in a Dynamic Society: a Modelling Problem. *Safety Science*, 273(2):183–213.

Richie, M. and Josephson, S. A. (2018). Quantifying Heuristic Bias: Anchoring, Availability, and Representativeness. *Teaching and Learning in Medicine*, 30(1):67–75.

Rogers, W. (1986). Space Shuttle Challenger Accident Commission Report. Available from: `https://www.govinfo.gov/content/pkg/GPO-CRPT-99hrpt1016/pdf/GPO-CRPT-99hrpt1016.pdf`. [Accessed 26 April 2021].

Rose, J. P. and Aspiras, O. (2020). "To hope was to expect": The impact of perspective taking and forecast type on wishful thinking. *Journal of Behavioral Decision Making*, (April 2019):1–16.

Royal Air Force (2019a). About the Tornado GR4. https://www.raf.mod.uk/aircraft/tornado-gr41/. [Accessed 21 January 2019].

Royal Air Force (2019b). About the Typhoon FGR4. https://www.raf.mod.uk/aircraft/typhoon-fgr4/. [Accessed 21 January 2019].

Rusk, D., Lin, K., Swartz, D., and Ridgeway, G. (2001). Bayesian updating of damage size probabilities for aircraft structural life-cycle management. In *19th AIAA Applied Aerodynamics Conference*, pages 1–11. Available from: https://arc.aiaa.org/doi/abs/10.2514/6.2001-1646 [Accessed 26 April 2021].

Salmon, P. M., Stanton, N. A., Lenné, M., Jenkins, D. P., Rafferty, L., and Walker, G. H. (2011). *Human Factors Methods and Accident Analysis*. Ashgate Publishing Ltd, Farnham, ebook edition.

Saunders, M., Lewis, P., and Thornhill, A. (2016). *Research Methods for Business Students*. Pearson Education, Harlow, 7th edition.

Savage, L. (1954). *The Foundations of Statistics*. Wiley, New York.

Short, J. and Sikora, I. (2016). Hazard Identification In Part-145 Approved Maintenance Organisations: The Management Of Risk Attitude. *City Research Online*, 37(9):1591–1601.

Shortridge, J., Aven, T., and Guikema, S. (2017). Risk assessment under deep uncertainty: A methodological comparison. *Reliability Engineering & System Safety*, 159:12–23.

Siegrist, M. and Sütterlin, B. (2014). Human and nature-caused hazards: the affect heuristic causes biased decisions. *Risk analysis : an official publication of the Society for Risk Analysis*, 34(8):1482–1494.

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1):99–118.

Simon, H. A. (1979). From Substantive to Procedural Rationality. In Hahn, F. and Hollis, M., editors, *Philosophy and Economic Theory*, pages 65–86. Oxford University Press, Oxford.

Sinnot-Armstrong, W. and Fogelin, R. (2015). *Understanding Arguments*. Cengage Learning, Stamford, USA, 9th edition.

Sloan, J. C. W. (2012). A400M wing assembly: Challenge of integrating composites. compositesworld.com/articles/a400m-wing-assembly-challenge-of-integrating-composites [Accessed 4 April 2021].

Slovic, P., Finucane,Melissa L., Peters, E., and MacGregor,Donald G. (2004). Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality. *Risk Analysis*, 24(2):311–322.

Society for Risk Analysis (2018). Core Subjects of Risk Analysis. Available from: `http://www.sra.org/resources` [Accessed 16 March 2021].

Spiegelhalter, D. J. and Riesch, H. (2011). Don't know, can't know: embracing deeper uncertainties when analysing risks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 369(1956):4730–4750.

Straub, D. (2009). Stochastic Modeling of Deterioration Processes through Dynamic Bayesian Networks. *Journal of Engineering Mechanics*, 135(10):1089–1099.

Straub, D. and Der Kiureghian, A. (2010). Bayesian Network Enhanced with Structural Reliability Methods: Methodology. *Journal of Engineering Mechanics*, 136(10):1248–1258.

Straub, D. and Papaioannou, I. (2015, 04014134). Bayesian Updating with Structural Reliability Methods. *Journal of Engineering Mechanics*, 141(3):1–34.

Swuste, P., Groeneweg, J., van Gulijk, C., Zwaard, W., Lemkowitz, S., and Oostendorp, Y. (2020). The future of safety science [in press]. *Safety Science*, 125(May 2020):1–9.

Tamasi, G. and Demichela, M. (2011). Risk assessment techniques for civil aviation security. *Reliability Engineering and System Safety*, 96(8):892–899.

Tang, H., Li, D., Li, J., and Xue, S. (2017). Epistemic uncertainty quantification in metal fatigue crack growth analysis using evidence theory. *International Journal of Fatigue*, 99:163–174.

Tavares, S. M. and de Castro, P. M. (2017). An overview of fatigue in aircraft structures. *Fatigue and Fracture of Engineering Materials and Structures*, 40(10):1510–1529.

The British Standards Institution (2018a). *BS ISO 31000 - 2018 BSI Standards Publication Risk management - Guidelines*. BSI Standards Limited 2018, 2nd edition.

The British Standards Institution (2018b). *ISO 19011:2018 - Guidelines for auditing management systems*. BSI Standards Limited 2018, 3rd edition.

Thekdi, S. A. and Aven, T. (2018). A methodology to evaluate risk for supporting decisions involving alignment with organizational values. *Reliability Engineering & System Safety*, 172(August 2017):84–93.

Theofanous, T. (2003). Risk Assessment and Management. In Milne, I., Ritchie, R., and Karihaloo, B., editors, *Comprehensive Structural Integrity*, chapter 1.19. Elsevier, Oxford, 1st edition.

Thöns, S. (2018). On the Value of Monitoring Information for the Structural Integrity and Risk Management. *Computer-Aided Civil and Infrastructure Engineering*, 33(1):79–94.

Thorne, K. M., Mattsson, B. J., Takekawa, J., Cummings, J., Crouse, D., Block, G., Bloom, V., Gerhart, M., Goldbeck, S., Huning, B., Sloop, C., Stewart, M., Taylor, K., and Valoppi, L. (2015). Collaborative decision-analytic framework to maximize resilience of tidal marshes to climate change. *Ecology and Society*, 20(1):30.

Todd, P. M. and Gigerenzer, G. (2012). What is Ecological Rationality. In *Ecological rationality: Intelligence in the world*, pages 3–30. Oxford University Press.

Torregosa, R. F. and Hu, W. (2013). Probabilistic risk analysis of fracture of aircraft structures using a Bayesian approach to update the distribution of the equivalent initial flaw sizes. *Fatigue and Fracture of Engineering Materials and Structures*, 36(11):1092–1101.

Toulmin, S. (1958). *The Uses of Argument*. Cambridge University Press, Cambridge, 1st edition.

Tran, T. B., Bastidas-Arteaga, E., and Schoefs, F. (2016). Improved Bayesian network configurations for probabilistic identification of degradation mechanisms: application to chloride ingress. *Structure and Infrastructure Engineering*, 12(9):1162–1176.

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157):1124–1131.

UK Ministry of Defence (2007). *DefStan 00-970: Design and Airworthiness Requirements for Service Aircraft*. UK Ministry of Defence, London, UK.

United States Air Force (2016). *MIL-STD-1530D - Department of Defense Standard Practice - Aircraft Structural Integrity Program (ASIP)*. US Department of Defense, Wright-Patterson AFB, OH, USA.

US Department of the Air Force (2019). *Air Force Instruction 90-802: Risk Management*. US Department of Defense, Washington DC, USA. Available from: `https://static.e-publishing.af.mil/production/1/af_se/publication/afi90-802/afi90-802.pdf` [Accessed 26 April 2021].

US Department of the Air Force (2020). *Air Force Instruction 21-101: Aircraft and Equipment Maintenance Management*. US Department of Defense, Washington DC, USA. Available from: `https://static.e-publishing.af.mil/production/1/af_a4/publication/afi21-101/afi21-101.pdf` [Accessed 26 April 2021].

Vagnoli, M., Remenyte-Prescott, R., and Andrews, J. (2018). Railway bridge structural health monitoring and fault detection: State-of-the-art methods and future challenges. *Structural Health Monitoring*, 17(4):971–1007.

van Asselt, M. B. and Renn, O. (2011). Risk governance. *Journal of Risk Research*, 14(4):431–449.

Vaughan, D. (1996). *The Challenger Launch Decision - Risky Technology, Culture, and Deviance at NASA*. The University of Chicago Press, Chicago.

Vermillion, S. D., Malak, R. J., Smallman, R., and Linsey, J. (2015). A Study on Outcome Framing and Risk Attitude in Engineering Decisions Under Uncertainty. *Journal of Mechanical Design*, 137(8):1–4.

Villa, V., Paltrinieri, N., Khan, F., and Cozzani, V. (2016). Towards dynamic risk analysis: A review of the risk assessment approach and its limitations in the chemical process industry. *Safety Science*, 89:77–93.

Virkler, D. A., Hillberry, B., and Goel, P. K. (1979). The statistical nature of fatigue crack propagation. *Journal of Engineering Materials and Technology*, 101(2):148–153.

von Neumann, J. and Morgenstern, O. (1947). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.

Wang, X., Rabiei, M., Hurtado, J., Modarres, M., and Hoffman, P. (2009). A probabilistic-based airframe integrity management model. *Reliability Engineering and System Safety*, 94(5):932–941.

Webster, J., Coleman, K., Mudge, A., Marquart, L., Gardner, G., Stankiewicz, M., Kirby, J., Vellacott, C., Horton-Breshears, M., and McClymont, A. (2011). Pressure ulcers: effectiveness of risk-assessment tools. A randomised controlled trial (the ULCER trial). *BMJ Quality & Safety*, 20(4):297–306.

Weiss, C. (2003). Scientific Uncertainty and Science-Based Precaution. *International Environmental Agreements: Politics, Law and Economics*, 3(2):137–166.

Wickens, C. D., Hollands, J. G., Banbury, S., and Parasuraman, R. (2013). *Engineering Psychology and Human Performance*. Psychology Press, New York, 4th edition.

Wilson, A. G. and Fronczyk, K. M. (2016). Bayesian Reliability: Combining Information. *Quality Engineering*, 29(1):119–129.

Wilson, J. (2010). *Essentials of Business Research*. SAGE Publications Ltd, London, 1st edition.

Woods, D. D. (2017). *Resilience Engineering*. CRC Press, London, 1st edition.

Wu, W. F. and Ni, C. C. (2007). Statistical aspects of some fatigue crack growth data. *Engineering Fracture Mechanics*, 74(18):2952–2963.

Yan, C. and Liu, K. (2011). Theory of economic life prediction and reliability assessment of aircraft structures. *Chinese Journal of Aeronautics*, 24(2):164–170.

Yang, J. and Manning, S. (1994). Aircraft Fleet Maintenance Based on Structural Reliability Analysis. *Journal of Aircraft*, 31(2):419–425.

Yang, X. and Haugen, S. (2015). Classification of risk to support decision-making in hazardous processes. *Safety Science*, 80:115–126.

Yang, X. and Haugen, S. (2016). Risk information for operational decision-making in the offshore oil and gas industry. *Safety Science*, 86:98–109.

Yasseri, S. F. and Mahani, R. B. (2016). Remaining Useful Life (RUL) of Corroding Pipelines. In *26th International Ocean and Polar Engineering Conference*, number July, pages 390–397, Rhodes, Greece.

Yates, F., Alattar, L., Eby, D. W., Molnar, L. J., LeBlanc, D., Gilbert, M., and St. Louis, R. (2011). An Analysis of Seatbelt Use Decision Making Among Part-Time Users. Technical Report UMTRI-2011-14, The University of Michigan Transportation Research Institute.

Yates, J. F. and Stone, E. (1992). Risk appraisal. In Yates, F. J., editor, *Risk Taking Behaviour*, pages 49–85. Wiley, Chichester, West Sussex.

Zarco-Macross (Wikidot) (2013). F-15 Eagle. `http://zarco-macross.wikidot.com/wiki:f-15-eagle` [Accessed 26 April 2021].

Zio, E. (2018). The future of risk assessment. *Reliability Engineering and System Safety*, 177(June 2017):176–190.

Appendix A

Project Gantt chart

ID	Stage Name	Sub task	Time								
			Year 1			Year 2			Year 3		
			Term 1	Term 2	Term 3	Term 1	Term 2	Term 3	Term 1	Term 2	Term 3
0	**Refining Research Proposal**										
	Continue Personal Academic Development										
1	**Literature Review**										
		Re-review early literature									
	Developing Research methods										
2	**Data collection/fieldwork**										
		Market research - ethical approval									
		Market research - civilian respondents - pilot interviews									
		Market research - military and civilian respondents - survey									
		Market research - results analysis									
		Paper drafting									
3	**Framework development**										
		Modes of usage									
		Process generation									
		Framework - draft test									
		Ethical approval - update									
4	**Framework Test**										
		Recruit candidates									
		Group testing									
5	**Results**										
		Develop new methodology proposal									
		Test new methodology proposal									
		Review new methodology									
6	**Writing the draft**										
7	**Final submission**										

= Original expected timescale

= Revised timescale

Appendix B

Participant recruitment marketing

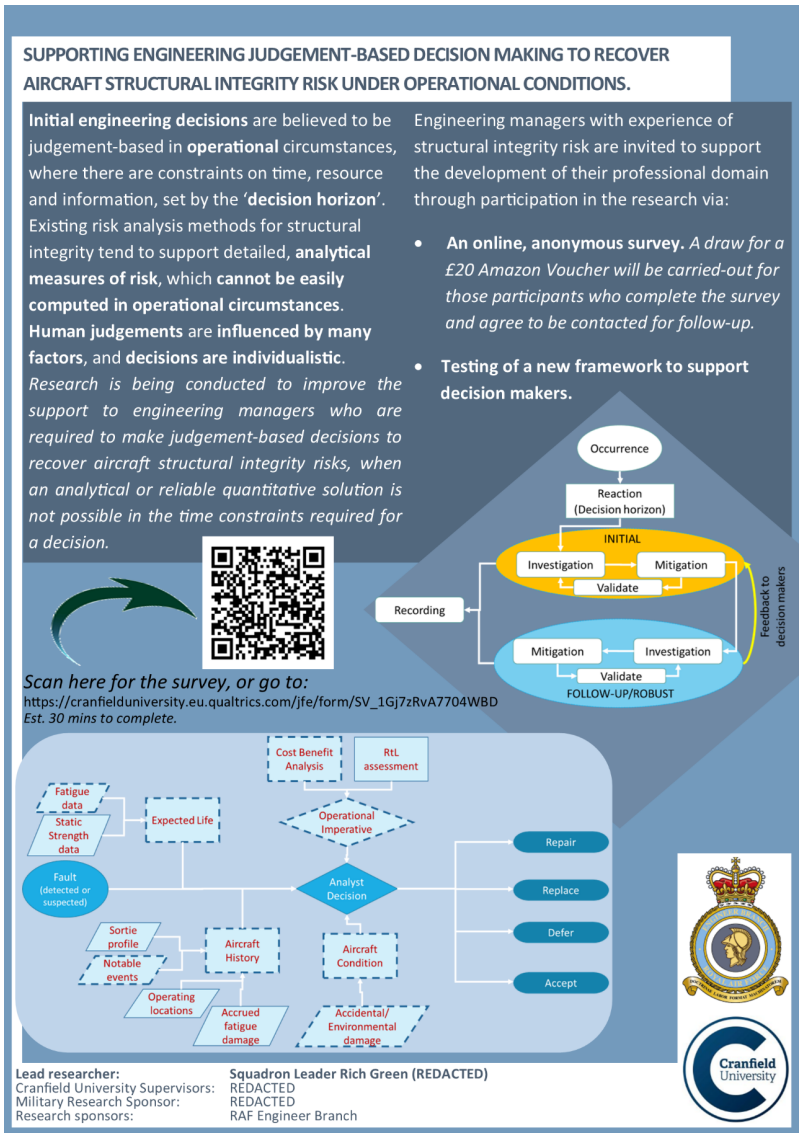


Figure B.1: Participant handout flyer distributed at internal UK MoD Royal Air Force Engineering Branch conference, 2018.



Figure B.2: Participant recruitment poster presented at internal UK MOD Royal Air Force Engineering Branch conference, 2018.

Appendix C

Interview transcripts

Pilot survey interview transcripts

Three of the seven pilot survey interviews carried-out during Phase 2 are transcribed in this appendix to provide an indication of the style, conduct by the interviewer, and responses given by the interviewees. Complete inclusion of all pilot interviews conducted does not enhance the readability of the thesis, but a sample is appropriate to illustrate the variation in responses.

Interviewee code	SIRM-120420190900
Organisation industry	Aviation in defence; company consults across multiple disciplines.
Organisation category	Operator/end user
Role when handling risk	Advisor and risk analyst

Interviewer: *[Discussion on the background of the research and motivation, and interviewee experiences.] So if you could just briefly talk about your industry that you're in and whether you class yourself as an end user, a policy maker, or none of those...and what your role within [company name] is.*

Interviewee: I'm part of a 3500 strong company. *[Company name]* is a UK separate trading entity which allows us to deal with UK defence work, and we have, our systems are independent from our corporate ownership, so we can hold secure data...We support high integrity industries, defence aviation and also land and maritime systems, civil aviation...and other high integrity industries would be medical, we do medical equipment assessment, bit of oil and gas, bit of rail, and so on. And so we provide consultancy support....we provide evidence based analysis which supports risk-based decision-making.

Interviewer: *So would you class yourself as advisors to the decision-makers then, or analysts?*

Interviewee: Bit of both, so we do analytical work, we produce artefacts, we do assurance so I'm the independent safety auditor on half a dozen different platforms for *[company name]*, so analysis, assurance, and certainly part of our role is advice on safety management and how enterprises can improve their safety management systems and their risk-based decision-making.

Interviewer: *So [company name] specifically is cross-industry, are you specifically aviation?*

Interviewee: All of my clients with the exception of one leisure company, which is less than 1% of my time, all of my clients are defence and most of them are defence aviation.

Interviewer: *That's all just contextual...Is risk analysis valued by your company? You obviously do your activity on behalf of other companies, so it's a difficult one to answer but I'm guessing by the nature of what you provide it's a valued activity?*

Interviewee: Yeh so we do invest in our professional development and competency management system, we invest in and recognise SQEP-ness *[Suitably Qualified and Experience Personnel]* in risk analysis techniques, fault tree analysis, loss models, quantitative risk assessment, and zonal hazard analysis and...

Interviewer: *So a variety of techniques?*

Interviewee: Yeh.

Interviewer: *In your experience, do you think the organisations that you support value risk analysis, or is it, they hire you in to do their work for them and then you go away and you might come back in a few years and it's not been touched, it's a time invariant solution that they see as a box ticking exercise?*

Interviewee: I think in seven years, the different organisations that I've worked with fall across that spectrum that you've described. There are clients with quite an immature view of safety management systems that just want an artefact that they can point to but aren't engaged in what it says. Most clients are engaged in some way, and so the safety products, advice and analysis that we're delivering is being challenged by....and our products are reviewed at safety meetings, so hazard review boards, safety panels and other integrity working groups, although we don't...ours is more of an assurance role in an integrity working group. Certainly our analysis is going into hazard review boards and safety panels. So there's a conversation in those meetings with LOAA-holding *[Letter of Airworthiness Authority]* crown servants about "what is this you're telling us, do we agree with it"...so tell me at what point it's useful to give examples...

Interviewer: *I'm specifically targeting structural integrity, have you got any examples of perhaps where....are people engaged specifically with structural integrity specifics, i.e. the peculiarities of structure, or are they more just trying to bring that along as another system that you can apply a QRA approach to and come out with a solution that looks like it fits the bill...I'm being cynical here, but...*

Interviewee: So most platforms have a loss model of some sort, might be a fault tree, which says the hazard risk for this platform can be modelled in the following way, and so that tool is being used to demonstrate whether the RA1230 aggregate design safety target is being met, and as you know you've got that hierarchy of accident, hazard and cause, and you can set the hazard at whatever level you like, I mean hazard is an artificial construct used by safety people to model risk... so you could have the hazard set to the accident level, but there's very few of them....or you could move further down that hierarchy and...I've seen platforms with 500 hazards. Generally we find that hazards work well when the platform risk is modelled around 40 to 80 hazards, and in there, some of those lend themselves to quantitative risk analysis using design...DO system safety assessments, and others don't, like structural integrity. There

will be a hazard that talks about failure of a SSI *[structural significant item]* for example, but the analysis can't be done, although I've seen it done in the same way, but often can't be done in the same way. And software, I would put in a similar category to structural integrity as it's deterministic isn't it, it's, you have a standard, if you meet that standard you're safe enough, so it's all about demonstrating compliance and if you can comply with *[DEFSTAN]* 00-970 then you're safe enough. I think programmable elements and structural integrity fit in that deterministic role, instead of the normal quantitative risk assessment that we can do with a lot of other systems. But they're all in there, structural integrity is in the loss model and unless someone, either consultants like *[company name]* or crown servants, unless someone challenges it and says we can't treat this in the same way, you can sleepwalk into thinking that QRA for structural integrity is a good thing and is a sensible approach to take, and I have seen that.

Interviewer: *So when you've got the specifics of structures, you mentioned "safe enough" and meeting DEFSTAN 00-970 to name one obviously, the MAA have got their specific ESVRE [refer to Section 2.1] format as well, and I see I'm targeting the recovering element of ESVRE, where you've got a designer who sets the design standard, you find a crack which is clearly not included in what the design was expecting to happen so you've got to do something about it...but you're still expected to show under the MAA regs a failure rate for the aircraft of* $1x10^{-6}$ *or whatever it be...in your experience, how do you find that you clients have justified the risk level that they're operating at, is it always a consistent threshold level that they have to show safety in accordance with, and when it comes to qualitative analysis is that threshold level appropriate?*

Interviewee: For structural integrity?

Interviewer: *Specifically for structures, yeh.*

Interviewee: I don't think there is much of a discussion about the probabilistic risk of structural integrity causing a risk to life. It tends to just sort of sit alongside. So all the fault tree and loss model you can aggregate up in a usual way, Boolean algebra gives you an aggregate platform risk, and compare that with *[RA]* 1230 safety target. I don't think, it depends, but structural integrity, if I were doing this now....I haven't worked on a fault tree or a loss model for a while....in my experience I think it just sits uncomfortably alongside. So you've got your system safety risk with an aggregate of....that often doesn't meet the design safety target, by the way, as many platforms bust the target as meet it, but and then you've got a structural integrity thing which sits alongside and if you're scoring yellows across ESVRE, then if you dig down in to *[DEFSTAN]* 00-970, then you do end up with some statements about "if you have compliance then your probabilistic risk is better than this, if you meet these design criteria"....that sort of sits alongside. The answer is it probably varies between platform...what was the question again?

Interviewer: *The threshold level that's set for safety, [RA]1230 sets that design target, do you think that's appropriate for structural integrity?*

Interviewee: I've not been in a conversation where structural integrity risk is discussed as the thing that threatens busting the design safety target. There are lots of other things that can...but....

Interviewer: *Is that driven by the fact that you can't generate a number easily, so it's not clear?*

Interviewee: Yep.

Interviewer: *And do you think that in generating that assessment, there is a lot of subjective judgement?*

Interviewee: In generating what assessment?

Interviewer: *The "structural integrity is safe somehow, yes there is a threshold level but we're not really discussing it in line with a threshold level because you can't"?*

Interviewee: Yeh that's right, it's just isn't discussed in the same context as the rest of the aggregate platform risk, or whatever words *[RA]* 1230 uses, so it just sits alongside as some ESVRE traffic lights.

Interviewer: *So what process is used then, they score against ESVRE and say how well they're meeting...well they're qualitative statements in RA5720 saying "for ESVRE compliance you have to do this, this, this"...so is it a qualitative score of "yes we're doing this" vice a "we've got a widget block on this part of the aircraft and it's got a crack in it and it's going to do this to the failure rate", which is what you'd do for systems, you'd have a black box that has failed and it's mean time between failure is this, therefore you can plot it's reliability and it's failure rate against the rest of the fleet so therefore how many would we expect on the rest of the fleet, but you can't necessarily do that for structure as we've just discussed, so to loop back round to the question...what's the method that they use to demonstrate that safe operating level?*

Interviewee: Yeh, so some platforms, and I'm going back years, I've seen some platforms years ago try and do that MTBF-type analysis on structural integrity erroneously and include it in the model. I haven't been involved in a fault tree for a long time and aggregating risk, but I don't think it is, I don't think that a number is put on it which is used in the fault tree...I don't think it contributes to the aggregate platform risk, held up against the design safety target, I think it sits alongside in an unquantified way, in a qualitative deterministic way.

Interviewer: *Do you think the decision-makers who are responsible and accountable for that safety of the platform, do you think they understand the why and the background to structure and why you can't make structure like that, clearly the assessments are going to have some form of judgement in it because it's qualitative, do you think they get everything, all of the background information, uncertainties, the assumptions that are made within assessments?*

Interviewee: I think the structures desk LOAA holder would be able to give a good account of why structural integrity doesn't feature in the fault tree in the same way as other systems. I think that understanding gets diluted once you get up to the safety panel level, but I still think there is an understanding of it, and I think most TAAs *[Type Airworthiness Authority[1]]* would say "yes I'm happy that we don't consider structural integrity in our loss model because over here I've got RA5720 series assurance, so I don't mind that we're not applying quantitative risk analysis on this, because I've got this other assurance over here and I've got my ISAA *[Independent Structural Airworthiness Advisor]* in the room and I've got the regulator in the room, and a whole room full..."....i mean it's a mature, those integrity working groups pre-date the regulator, don't they. They pre-date 2010, particularly structural integrity, which

[1] A delegated airworthiness holder for an aircraft type in the UK Ministry of Defence.

goes back probably 30 years since Buccaneer failures or something. So these are mature assurance processes and the TAA can be confident in them. So I think they're treated separately and independently although clearly it's the same guy chairing the structural integrity working group is also chairing the safety panel, so that safety panel is fed by the hazard review board, quantitative risk and the SIWG *[Structural Integrity Working Group]* deterministic compliance and assurance, and a safety panel says "yeh I've got my systems aggregate risk, this is my structures compliance, I get that that doesn't really feed in to whether I meet my design safety target or not, but that's ok, because actually to do that is either too difficult or probably erroneous". So I think there is an implicit recognition that these two things are different.

Interviewer: *Most of our discussion there was concerning military clients, have you had any experience of similar aspects with non-military clients?*

Interviewee: No, I haven't got much experience, although my crude understanding of it would be that commercial operators don't spend much time managing risk, it's all about design certification because they are dealing with a number of aircraft which are very...their configuration very closely aligns to a global fleet so it's all about design certification to EASA or FAA design requirements and then they're buying a design compliant platform and don't spend much time managing risk.

Interviewer: *In your experience, and to naturally follow on from that, who do you think uses the risk analysis and the decisions that come out of the risk analysis? Do you think it's always operators, or manufacturers?*

Interviewee: Are you back in defence now?

Interviewer: *Wherever, across the board, if it's different across them then it would be useful to know where it's different and how different industries use that risk analysis for structure specifically that is...*

Interviewee: Ok, structural integrity is not usually expressed as a risk, it's expressed as assurance and compliance.

Interviewer: *Do you think it's proactive then, or reactive? So you've got a whole load of systems that are in place and allow you to declare that the platform is safe because you designed it in accordance with this...do you think that's a proactive approach, or mostly they'll react when something happens?*

Interviewee: I think there is relatively little communication of structural integrity risk to the holders of the risk, so I think there is very good communication of risk to the holders of the risk so what do I mean, I mean that the Type Airworthiness Authority if I stick with defence aviation, and I'm the ISAA for [list of aircraft types], so across these types there is very good communication of risk and this is improved and continues to improve over recent years...very clear expression and communication of risk to the duty holder, and the emerging RA1205 requirement for duty holders to have their own air system safety case will only improve that the way that is done. Very little of that is to do with structural integrity risk and that's probably ok because the risk in my view is relatively low because the DEFSTAN for structural integrity design has been around for decades and is well understood, and the SIWG governance and assurance has been around for years which pre-date the regulator and is well understood. And so all the way from the DO and [other company name], the delivery teams, industry and regulator all understand

structural integrity and there's a good assurance process there, and if there were a structural integrity risk it would be highlighted in the right way. So in fact, if I think about rotary wing, I think about structural integrity risks, that I'd have to look up some details, but I can think of structural integrity risks to rotary wing which are communicated to the duty holder in an appropriate way, when they spike up above "yep everything's ok", then there is a good healthy dialogue about that risk, and is expressed and communicated in the right way.

Interviewer: *That sounds very much steady state, what about when something emerges that's unexpected, and you enter into the "recovering" side of ESVRE?*

Interviewee: Yeh, again I think...what, about the way in which that's communicated?

Interviewer: *Communicated, managed, because the assurance and design in accordance with DEFSTAN 00-970 is great until something goes wrong that you're not expecting and all of your design assumptions then are challenged by "oh we've had something that has happened that has potentially thrown our expected safety standard for the aircraft in to doubt" [example given from UK military aircraft]...so your entire design case was challenged by this emergence, so I'm interested in whether that is a common thing or whether it's something that you think is fairly isolated and doesn't happen that often...*

Interviewee: Ok, I mean I'm not close to that any more, if I look back over years then I would say that structural integrity risk was always, you're asking about how well it's managed right? When you get an unusual....

Interviewer: *reactive, something that you have to react to*

Interviewee: I mean that's always been done really well, it's always been communicated to, you know we used to have those chief engineer notices, annex As and annex Bs, there was always good communication and you had to have your safety arguments robust because they were being challenged up to at least 2-star level, and there were plenty of mature corrective action options, special instructions technical, followed by modifications, all these things are well understood and I see less of it now, but when I see it, I see the same behaviours and similar processes being used to manage structural integrity risk. I don't think we've lost that, what has happened clearly in the last 10 years has been far more emphasis on QRA and use of hazard risk matrices, use of loss models, RA1230 compliance, but I don't think we've thrown away all that...some of what I was talking about there pre-dates all the recent systems safety work, I don't think defence has forgotten how to do that. I think it's as robust as ever.

Interviewer: *Do you think it runs contrary to the requirements of [RA]1205, or [RA]1230, with the design safety target, which is a quantitative value, there's nothing in there that says "you have to be safe enough from a structures point of view, you have to have your risk of loss life to be less than this threshold value, and your tolerable and ALARP upside down triangle has very clear thresholds where you're allowed to operate but you can't have those thresholds. It sounds like what you're saying is "it's safe because we've got a robust and well resourced way of managing it, but it's not necessarily in-line with what policy would like us to do".*

Interviewee: No it's not explicitly linked. The structural integrity policy is not linked to the RA1230 design safety target in an explicit way. I think you could create that argument. If you reach deep enough into *[DEFSTAN]* 00-970, you come to chapters with S-N curves, which talk about "if you're compliant with all of these design safety factors..."

Interviewer: *Scatter factors?*

Interviewee: Yeh, I remember factors of 3 and factors of 5, but can't remember how they're used. But if you comply with those then your probabilistic....I mean there are probabilistic risks buried deep in *[DEFSTAN]* 00-970 that say "if you do it in this way with the right safety factors, your probabilistic risk will be better than..." whatever that number is, which I think is probably coherent with the design safety target, but I don't think anyone expresses it in that way. I think it all holds together, but no-one's making the direct link.

Interviewer: *...Do you think in that sense structural integrity risks are handled conservatively?*

Interviewee: Differently. Because it's so hard to express structural integrity risk in a probabilistic way is why we have an ISAA in the room at SIWGs, and TAAs know enough that when the ISAA says "I'm concerned" they need to pay attention, but equally...most structural integrity risks...if there's a route map to improve it and the ISAA holds the meeting to account "how are you doing with this"...the delivery isn't allowed to forget about the corrective actions that they have to do through modification or inspection to keep safe. It's just that that risk is probably never expressed probabilistically. It's just managed differently. Is it more conservative? No, in many ways I think there's a pretty grown up treatment of structural integrity risk because SIWGs are mature and there's a pretty good level of expertise in the room, isn't there? You've got DO SQEP, you've got the ISAA SQEP, as well as all the experience of the joint community in the room, so I'm not sure about conservative, I think it is informed risk-based decision-making, but it's not quantitative risk analysis. These are rather personal opinions that I'm expressing.

Interviewer: *But your personal opinion will be formed on the back of your professional experience.*

Interviewee: Yeh.

Interviewer: *One of the reasons I anonymise the data is because I need personal opinions on this to be able to get around any organisational bias which might creep in. Do you think on the back of that, it's easy to describe a consequence and a likelihood for structures?*

Interviewee: No that's not....that's seldom expressed. Everyone understands the consequence is bad...I think straight away there's an implicit understanding that this is a catastrophic...if you try and put this in a hazard risk matrix it's going to be in the catastrophic column, they'll think severity...assuming it's an SSI then severity is catastrophic, probability is not expressed for these risks because it probably can't be.

Interviewer: *And yet it's supposed to be, because you've got this hazard risk matrix which your decision-maker wants everything to be expressed in so they can compare across them. That's one of the reasons why I'm exploring this is because decision-makers who want everything in a single meeting, in a single format and it's driving, pushing structures into that format. To what extent do you see subjective judgement used in deciding upon structures?*

Interviewee: So it's not quantitative, so all risk-based decision-making is qualitative. But it's not subjective...it is evidence based, but the evidence is compliance, is deterministic evidence about compliance with policy and process, rather than probabilistic failure rate.

Interviewer: *Do you think in an operationally urgent situation, so time and resource critical, where you perhaps haven't got....I'm thinking here where you've got maybe a deployed aircraft....a type of circumstance...a deployed aircraft, something happens...do you think in those circumstances it's easy to justify a decision regarding a structural integrity risk? Or is it the sort of decision that necessitates having a lot of expertise around before you come to a satisfactory conclusion?*

Interviewee: It's just a different sort of evidence, your safety argument, if you were a duty holder of a TAA in justifying ALARP-ness in a court of law, then you'd just say "I had a room full of SQEP when I made this judgement, I had the right SQEP in the room, we made this decision".

Interviewer: *Is that decision often written down, like captured in terms of.....or is it more of a case of "following a discussion we talked about this, this and this, and we therefore decided this", is it mostly text based? Is there any kind of argument put together in a coherent sense? Or is it very much a case of "the conversation we had led to..."? Is there a clear audit trail basically?*

Interviewee: Yes, again the minutes of the *[RA]* 5720 series meetings are robust and supported by a strong slide pack. But of course there are requirements now, which were duty holder advice notes [DHAN] in old money, but are now called SEINs (significant engineering information notice) and there's a *[other company name]* process which says how TAAs communicate risk to the duty holder and I've seen those used for lots of different risk communication to the duty holder. I can't point to one that's been used for structural integrity but that may not be...I wouldn't like to say the normal risk communication process doesn't apply to structural integrity, I just haven't got a ....

Interviewer: *Ok, what changes should be made to improve or change organisational handling of structural integrity risk?*

Interviewee: I don't think that it's far off. I think ESVRE assessments are reported to the 2-star ESR, so the 2-star is saying "how are we doing in our integrity management?" You're talking about risks perhaps which then are out of process and the things that spike up, which fall outside of *[DEFSTAN]* 00-970 compliance and I just...I'm confident in and reassured by the increasing emphasis on characterisation and communication of risk from the TAA to the duty holder. The DHAN, now this SEIN process, and the increasing emphasis on the duty holder to have an ASSC *[air system safety case]*, everything I've seen is safety positive in the way risk is communicated to the....is captured and communicated. I think that structural integrity risks would be handled in that way regardless of the fact that they sit alongside the loss model, they don't fit the loss model in the normal, they don't contribute to *[RA]* 1230 in the normal way. I don't understand software risk really, or programmable elements, but I've got this suspicion that there's a similar sort of conversation to the one we're having now, that you could have with a programmable elements guy that says "how do you...you know you can't quantify the risk in the same way, so there's..."

Interviewer: *Well you only have to look at the 737MAX issue to understand the issues with it there.*

Interviewee: Yeh, so I think it's just baking into.....I've got it somewhere on the system....making sure that all these safety meetings, safety panels, for example are...and the duty holder facing risk communication process, make sure that those are explicitly required by regulation to consider programmable elements and structural integrity risks, in addition to systems risk, which can be expressed in quantitative terms...I think all of the processes are there, we've just got to prompt everybody through regulatory requirements at every step to say "make sure when you have a safety panel...when you have a hazard review board...when you produce a DH facing expression of risk...don't forget to express the structural integrity and PE *[programmable elements]* risk" because they're different, but they've got to be considered at the same time. I suspect that probably happens, but I'm not sure how well it's mandated as a requirement.

Interviewer: *What information do you think is essential in order to make a decision upon a structures risk?*

Interviewee: An explicit judgement about, by a room full of SQEP, about whether the risk breaches the design safety target.

Interviewer: *Is there anything specific that they might need, in terms of fleet hours, identifying peculiarities of the exposure, or the traits of a particular issue like crack propagation rate or something like that, or in your experience do they not go down to that level to make that decision?*

Interviewee: No I think they always go to that level. There's always that discussion of metallurgical, you know have we got fracture surface analysis, what do the metallurgists think?....what do 1710 NAS[2] think? Yeh absolutely crack propagation rate, inspection interval, confidence that you're going to... what type of crack is it, are you likely to find it? That's always bread and butter discussion of a structural integrity issue, but when you've had all that discussion "so, we've got some maintenance mitigations here of inspection, we've got some operating mitigations here, g-limit or something, and we've got a modification coming along the line so we can claim ALARP"....but with all of those the room of SQEP needs to form an explicit judgement about have we breached the design safety target. At the end of a safety panel, the TAA is making explicit assertions, which are required by the operating centre safety management systems, but there's always a slide at the end of a safety panel that says "as the TAA, I'm happy that all hazards are tolerable and ALARP" and I think an explicit...ok, so what's my role? When I sit in a SIWG, I'm thinking about the linkage between integrity working groups and the hazard log, and the question I ask the most in a SIWG is "is this cause something which is a new cause in the hazard log?" and of course, actually, you could still have a structural integrity risk in the hazard log without quantitative to do with expressing it. I think some kind of explicit safety statement...SIWGs...you get 150 slides and there's lots of useful information on them, but the safety statements that associate with them are often implicit. You'll never see a slide which says "this structural integrity risk is tolerably safe because we are...." you always need a summary slide which says "this SSI failure is

[2][A UK Ministry of Defence organisation specialist in structural assessment and repair.]

tolerably safe because we've increased the inspection interval, we've changed the SOIU *[statement of operating intent and usage]* and it's ALARP because we've got a mod on the way". These things are never stated in safety terms, so I'd like to see safety case-type claims made, explicitly on SIWG slides, and I'd like to see a TAA make a safety statement at the end of a SIWG, which says "there are no new causes here" or "the risks associated with the hazards we've discussed are considered as..." however that's expressed...not breaching the design safety target. I don't know how easy that is to do.

Interviewer: *In the case where you said "there are no new causes", those causes obviously have to be presented by somebody, do you think there's a threshold level where some low level risks don't get presented? Or is that not something that you've necessarily had sight of in your role?*

Interviewee: Yeh, in a way...I can't think of an example of a structural integrity risk that hasn't been....I mean all structural integrity risks will have been assessed by an LOAA holder in some way.

Interviewer: *But do they all get presented at the SIWG, so the decision-maker decides that "this is contributing, this is tolerable and ALARP" but the LOAA decides what goes on the slide pack.*

Interviewee: Yeh, but it's pretty...they'll present all the SITs *[special instructions technical]*, they'll present all of the 760s[3] with a structural integrity twist on it, but yeh I'd say it's complete.

Interviewer: *Do you think there are any uncertainties or assumptions that decision-makers don't understand about structural integrity aspects that perhaps they rely almost entirely upon their advisors to say "this is good, therefore you don't need to worry about it".*

Interviewee: Yeh, I think the safety cases don't say much about the fact that structural integrity is managed differently, and the fact that it's hard to express it quantitatively. I think safety case reports could express the nature of structural integrity risk in a way that would just remind people that it's different...would capture the sort of points that we've been talking about.

Interviewer: *So you think therefore that decision-makers don't either communicate it, or don't get it? Do you think there's assumptions in the background that perhaps they don't understand...a great example would be Tornado where the assumption is that the design usage is almost entirely altered, so when you come to do a life extension you assume the platform was operating at mid level with...not flying over the German plains any more pulling high-g turns, and those kind of assumptions.*

Interviewee: No I think all that...the SOIU is, again across a number of platforms, I see a mature and robust review annual and triennial review of SOIUs. I think that's being done effectively. But I think there are...some of the understanding about the treatment of structural integrity risk is implicit. Yeh there are assumptions they're about how it all fits together, and the judgements being made are the right ones, but often I think there's quite a lot that isn't spelled out, so we're relying on implicit assumptions by good experienced people rather than...and I'd prefer to see more, especially as demographic and other influences mean that the make-up of [other company

[3] A UK Ministry of Defence aviation form reference, used to request amendment to published aircraft maintenance information.

name] changes, the way minutes are written, I've seen...has deteriorated appallingly. The way minutes are being written is really poor compared to the way, the old school Service-writing way of doing things, and that risks losing explicit expression of the basis for judgements. So I think it's all ok, but the way in which we express these judgements is in danger of being lost.

Interviewer: *That's really encouraging because that's exactly what I'm trying to target, it's the communication of that judgement and how that judgement is recorded at a given time...*

Interviewee: I've got a lot of respect for civil servants, and there a lot of young, keen and ambitious civil servants coming through but they can't write. The problem with structural integrity is, because it's not expressed in a...there are hazard evaluation forms where you document very clearly the basis for severity and probabilistic judgements, because structural integrity sits outside that you've got to capture the basis for decision-making in a different narrative and these minutes of SIWGs...I think I'm contradicting what I said earlier, I said SIWGs were mature and effective earlier, but I am going to contradict what I said earlier because in some cases, while the process is mature, the quality of the minutes, which document the basis for decision-making are increasingly flaky.

Interviewer: *So the intent is good, but the execution is lacking?*

Interviewee: Yeh, that's right. The regulation and processes are good, but the quality of the narrative evidence has deteriorated significantly in the last decade or more.

Interviewer: *[Discussion about a question that would be sent to the participant following the interview, regarding risk attitude.]*

Interviewee: The sort of things that came to mind when you were saying that were it's less and less [other company name] and LOAA holders that take a view on risk, or have to accept risk. The behaviour is increasingly [other company name] has to express risk which it then transfers for a duty holder to decide, so it's about expression of risk, but the duty holder is the decider. And my experience of duty holder risk appetite is that, and I've had one-to-ones with two or three ODHs on specific risk items, and ODHs generally have a high risk appetite...they're not bothered about busting $1x10^{-6}$ safety target if there's an operational imperative, with the one caveat that there has to be a story about how this gets better. There has to be that waterfall diagram, where the TAA says "right you're out of process now, we've bust this safety target, tell me when the fleet will be back in process and I'm going to hold you to account on that". But if it's a temporary, you know something's happened, ODHs in particular will keep flying provided there's a convincing story of how it gets better. But perhaps that wasn't the question you were asking.

Interviewer: *Do you have any last comments, I know you mentioned that point there, but is there anything else you wanted to mention that you think's relevant?*

Interviewee: No, I've always wondered about whether, as ESVRE scores go from green to red, whether there should be some sort of equivalent where you say "when you've got an amber, you should manage this as if it was an order of magnitude that's the wrong side of the design safety target. If it was a red, you should manage it as though it was two orders of magnitude the wrong side of the design safety target", but it's putting arbitrary numbers on non-compliance.

Interviewer: *The accompanying question to that would also be: is there a criteria for them declaring red, amber or green?*

Interviewee: Well, yes, the ESVRE, the SARCS scorecard tells you what the traffic lights, if it needs a 2-star to fix it, then it has to be amber, so the criteria do exist for traffic lighting ESVRE.

Interviewer: *I never had sight of the SARCS score...thank you very much for your time, sorry for running over...*

Interviewee: Stepping back from everything and how would I summarise everything that I've just said....the safety management techniques and discipline, quantitative risk doesn't have much to offer, but the use of the clear expression of risk in equipment safety assessments and in air system safety cases, there is some good rigour there that could be brought to the structural integrity management governance and assurance, and some of those safety claims could be tied into the structural integrity management in a more explicit way. There won't be quantitative risk analysis, there'll be qualitative, they'll be done in their own way, but you can still make some safety claims in the safety case, which are tied in to the SIWG itself. So the SIWG knows, it's sign posted as it goes through how it contributes to the safety case.

Interviewer: *I'm very grateful for your perspectives on it, thank you so much.*

Interviewee code	SIRM-270220191400
Organisation industry	Aviation
Organisation category	Designer/equipment manufacturer
Role when handling risk	Advisor and risk analyst

Interviewer: *[Discussion about the background to the research.]*

Interviewee: I am from the background of a fatigue and damage tolerance engineer. So I've got through all the deterministic stuff, and now I'm in a position called a technical authority. So really we're sort of representing EASA in our case, in *[company name]*.

Interviewer: *So are you a [company name] employee or regulator?*

Interviewee: Absolutely *[company name]* employee but as our design office approval, we're expected to follow certain processes in order to keep that approval.

Interviewer: *And you're essentially an independent authority?*

Interviewee: Yeh, for instance this morning I was sitting on a review panel as an independent person, but I'm not independent in the strictest sense of the word, but I am able then to put an independent view on everything

Interviewer: *Like an internal QA sort of thing for structures?*

Interviewee: Yeh, and consequently I have the authorisation signatures to let everything go out, anything fatigue and damage tolerance related to wings I have to sign.

Interviewer: *What level of risk do you think you hold relative to the organisation?*

Interviewee: In terms of personal risk to me?

Interviewer: *Well organisation, so I used to hold authorisation to be able to approve deviations for Tornado for the fleet, but only for what we class as low risk, so we had a risk matrix of likelihood/consequence and if you landed in the low area then that was mine.*

Interviewee: I authorise all the maintenance programs for wings.

Interviewer: *So any level up to high level of risk, could be catastrophic for the fleet, type thing?*

Interviewee: Yeh.

Interviewer: *So, from what [a colleague of the interviewee] told me, it sounded like, because of the way the civilian regulations operate, they don't allow you to operate with damage, any time you find damage you have to correct is straight away.*

Interviewee: That's correct.

Interviewer: *Where I'm coming at this from is an in-service arising you have to do something about it, in the military you're allowed to operate, but that level of risk is taken by people like myself where you could go to the OEM and they could go "yeh we'll tell you how long you can live with that crack for, you're going to have to repair it at some point but if you operate under these assumptions you could fly for another 50 flying hours."*

Interviewee: Just to talk a bit around that, we will do ferry flights, so I'll get involved in the decisions for those. So if we have damaged structure...

Interviewer: *To get it to a suitable maintenance?*

Interviewee: ...a non-revenue ferry flight as we would refer to it

Interviewer: *...With no fee paying passengers?*

Interviewee: Absolutely.

Interviewer: *You're keeping the risk at a minimum?*

Interviewee: One or two flights at the most and usually operating restrictions on the aircraft. The FAA have a set of rules in their make-up to allow flying with damage and essentially you have to maintain the original capability of the airframe, so it has to be ultimate capable.

Interviewer: *So you have to still be able to reach ultimate limit load with that damage in-situ?*

Interviewee: With no progression of it in that period.

Interviewer: *Yeh, admittedly your exposure is low because it's a small period of time and it's one flight...*

Interviewee: Well no, but you can potentially extend that. The ferry flight thing can be done as just a discussion with the authorities. The flying with damage is something that's kind of written into guidelines that the FAA have produced but they are quite harsh guidelines so you do have to show ultimate capability.

Interviewer: *How do you go about doing that?*

Interviewee: Well you have to show ultimate capability with the crack in place, so that is a deterministic analysis.

Interviewer: *Which, as I understand having been told about this the other day and dug into the FAA regs, it has to be backed up by empirical or test evidence*

Interviewee: It's next to impossible.

Interviewer: *So how are you supposed to do that in such a short period of time?*

Interviewee: We would be able to, because we would probably be looking at this for a fleet management situation, because we have a slightly different issue in civil...grounding one aircraft to deal with it isn't so much of a problem; the problem is when you're trying to deal with an issue that could be affecting the fleet. So these kind of topics kind of come up for fleet management of several hundred aircraft.

Interviewer: *"We've got a crack on this aircraft and it's done these flying hours and landings, are we going to have it on these?".... That's obviously got to be a fairly quick turnaround on a decision?*

Interviewee: It doesn't necessarily, if you can show that there's no unsafe condition and that you can continue operating the aircraft...

Interviewer: *Could you just expand on unsafe conditions?*

Interviewee: Unsafe condition...so you're not falling below the levels of safety that you expect according to the regulations.

Interviewer: *How do you know that you're not falling below that with a crack of a possible indeterminate length?*

Interviewee: Well you have to show that you'd be able to find it before it before it became a dangerous length. So a normal damage tolerance philosophy would show that you have to be able to have limit capability if you had a crack occur and you have to be able to find it....or have a suitable opportunity to find it before it gets there. So if you had the capability of a large damage, then you'd have to show that you can find that, within that type.

Interviewer: *[Discussion about the military context, and safe life philosophy]*

Interviewee: Well we do have safe life structure and we have had to do things like this on safe life structure...sometimes that's done qualitatively, sometimes it's done a little bit by experience in terms of managing the risk. If you know that a part of the aircraft is inspected regularly, I know that a lot of military regimes are quite regular inspections

Interviewer: *Can be, it depends on the platform.*

Interviewee: Then you could potentially discuss with the authority but we would be discussing this with the authority in this case, so we would have to talk to EASA about this, it wouldn't be entirely within our decision to make.

Interviewer: *You'd have to be transparent?*

Interviewee: Well we would be making the decision, but they'd have to agree it.

Interviewer: *I'm interested in the ones where you were saying sometimes there's an experience-based element to it, so how much risk do you think ....there's a human subjective element...when you speak to someone who's a fracture mechanics expert that's their day job, they'll be saying I've got base assumptions that I need to operate from to be able to say if a material's going to fail or not. Now if you're operating in an experience world, are you basing that on "hey we've seen a crack like that before in exactly this type of material and it was fine.." is that the kind of judgement...*

Interviewee: Yeh and it will be things like looking at which aircraft you found it on, how does it sit in the fleet.

Interviewer: *Looking at the life characteristics as well?*

Interviewee: Exactly, so has it happened out of characteristic with the rest of the fleet, or is it the fleet leader? If you find damage on the fleet leader....you've potentially got the tip of the iceberg and in order to try and gain the experience from that you'd probably instigate some fleet leader inspections which could give you the confidence that next aircraft around it are undamaged. Therefore you can get the confidence that actually, we need to keep looking at them, but we've now got a little bit of experience that it isn't every aircraft. If it's an aircraft which is somewhere in the middle of your fleet then you may have more confidence that it's a one-off.

Interviewer: *So that's quite a qualitative risk assessment for the fleet, is that something that the regulators are comfortable with you operating in that sense?*

Interviewee: Yes, for short periods of time the expectation is that we would then do a full route cause analysis and establish if there's something special about that aircraft. If there is, then you can effectively take it out the dataset. If there's nothing special about it then we would carry on with a more conservative approach to check, to statistically show that it's not special, or is special, depending on which way it goes. So there is judgement in this and some of it can be quite intangible judgement...

Interviewer: *Is the requirement for judgement driven by restrictions on time and resource, so you're under pressure?*

Interviewee: Not necessarily, you've got the difficulties where you don't have the information and you don't have any way of getting the information because we don't have the link into the fleet.

Interviewer: *So you're talking about lifing information and usage statistics.*

Interviewee: Not necessarily that, but we can't go and look at aircraft, we can't just go and do it.

Interviewer: *So you're relying on the measurements that they've given you?*

Interviewee: Possibly, or we would have to form a relationship with an airline, and then do inspections through that. We can't just go and look at the aircraft, so actually gaining the information back from the in-service aircraft can be quite difficult.

Interviewer: *I'm guessing that's in their interest though?*

Interviewee: Their interest is flying.

Interviewer: *But if you've got a crack and they're saying we want to continue operating?*

Interviewee: Yeh, obviously they have the safety side of it as well, and that's where the airworthiness authorities would come in to mandate inspections if necessary but they're not our fleet...we very much sold them and they've gone out into service, so that's probably where it's slightly different for you

Interviewer: *Yeh quite possibly, because we were effectively the operator and the decision-maker on what happened. Would you consider yourself a decision-maker, or risk analyst, or advisor, or commentator?*

Interviewee: If you look at the chain of command in terms of who's making the decision, it's probably not me, I'd be very close as an advisor, I could be the one that makes the decision that's accepted.

Interviewer: *So you might be an analyst-type role then? An advisor is not someone I'd consider who would take the fall if it all went Pete Tong.*

Interviewee: Yeh I wouldn't take the fall because it ends with the chief engineers, but they would look to me. So they're the decision makers but they would expect me to give them the answer.

Interviewer: *Do they always understand the assumptions and uncertainties that are in structural integrity analyses?*

Interviewee: It's my job to ensure they understand the right ones, so I need to manage that part of it.

Interviewer: *So you have some that you wouldn't tell them?*

Interviewee: You wouldn't confuse them. You'd keep the information as necessary for them to know.

Interviewer: *What sort of assumptions and uncertainties would you tell them, like usage and lifing?*

Interviewee: I would aim to explain why I'd come to a decision because I'm sitting in a role between the analysts, here. So I would ask them to do certain analysis, and they would do that, and then there's the test information and the background of other programmes and things like that, I bring that information together and I would form a rationale.

Interviewer: *Sounds like a qualitative judgement?*

Interviewee: Yes it is, very definitely qualitative, and I would say based on this, this and this, I've decided this. And that's the kind of information I would say to chief engineers, and they may then discuss with me "what do you feel the risks are with various parts of it" and I might be able to say "this is more questionable than some of that, but we've got this as a back up", so I can also sometimes help with the strategies for dealing with a problem as well.

Interviewer: *Have you ever seen any situations where a decision makers has made a decision and perhaps either not understood everything, or has maybe later gone on to go "we didn't know that when we made that decision at the time" so 20/20 hindsight-type decision?*

Interviewee: There are certainly decisions that we've made that we haven't known everything and we would've done something different...had we known. But you couldn't have that hindsight....you couldn't have got to the position where you would have known that.

Interviewer: *Of course because of the information you had at the time... [discussion about military aviation maintenance circumstances.]*

Interviewee: When discussing with the authorities, because I'm involved also with the discussions with the authorities directly, we would never use a probability, for example.

Interviewer: *So you would never quantify a likelihood?*

Interviewee: I don't think so, I think we would use things like "likely", "unlikely", "probable"

Interviewer: *Do you attach any value to those?*

Interviewee: Not officially, but there's so many different things that would come into it. Some of that's fleet size.

Interviewer: *When you say fleet size is an issue, is that because statistically you might not have enough information?*

Interviewee: Fleet size can work in several ways. If you've got a big fleet then potentially flying for a month with a big fleet you're exposed to more risk than you are flying with a small fleet for six months so you know we've got tens of thousands of aircraft. So if we were to find something that affected all our aircraft, that's quite hard to justify because you've got such a big fleet.

Interviewer: *But you also might have rare events on a small fleet that ...it might be a rare event that becomes a common event but it's happened on one airframe and it could affect all of them but you haven't got enough evidence to say "this is a rare event", or whether it's....*

Interviewee: We had an event where we had a one-off, as it was called at the time, and we checked it could be safely found without exposing the aircraft to very long in a loss of full capability. So the event would cause a loss of capability but only for a very short duration. So the argument there being you're losing full strength of the aircraft but for a very short time and you would safely find it. There's no analysis in there, and the judgement was based purely on a feeling for what felt like a reasonable interval.

Interviewer: *[Discussion about research aims being aligned to this scenario]*

Interviewee: We often try and link back into the maintenance tasks. In that particular example there was a maintenance operation...it was greasing and it was done regularly. Actually the examples that we found had been found during that operation so we had a good confidence that that operation would find the damage. So we had this backup that this operation happens regularly, the greasing, so we're going to catch it. We did also back this up with some deterministic analysis.....

Interviewer: *Assuming the maintainers are good and vigilant.*

Interviewee: Yeh exactly, but at the time remember it was a one-off. A second event occurred and it was found during the same operation, and that continued pretty much. We improved the way we inspected the aircraft, a lot of that was saying "when you're doing this operation, please look at this" so again you're managing the risk of someone doing the greasing operation while looking the other way.

Interviewer: *Do you do that through service bulletins?*

Interviewee: Yes, normally, I think that one was.

Interviewer: *[Discussion comparing with the military technical instruction.]*

Interviewee: This may have been a directive actually, so it would've been ...

Interviewer: *An airworthiness directive?*

Interviewee: Yeh, that's a "you have to do this".

Interviewer: *Are there optional ones?*

Interviewee: Well we can't mandate anything, which is an interesting side of things. So we can't make anything mandated, that can only be done by the airworthiness authorities.

Interviewer: *So the ADs come out of the CAA, or whoever?*

Interviewee: Yeh, or EASA in our case. So we would recommend a certain course of action and they decide whether it's going to be recommend to the airlines, in which case, don't make a choice based on some criteria which they decide, or whether it's actually mandated in which case they have to be followed by everyone operating in EASA airspace or under EASA conditions, and FAA as well because they sort of share it.

Interviewer: *[Discussion about similarities between civilian and military domains.] Do you have any techniques where you proactively look for risks, structural risks, unknown unknowns....?*

Interviewee: I like looking for damage capability. So when I'm reviewing designs, which is something I have to do regularly, I'm not so much looking for how someone's analysed the crack growth as the damage capability though it's useful information, I'm looking to see if that breaks, where's the load going, what's going to break next, how are we going to find. And by making our designs follow that principle...it's really trying to get the damage tolerance principle into the aircraft as much as possible....so by following that you do build in risk management and it allows you to deal with a problem that you don't know is going to happen.

Interviewer: *So effectively you're management of that is by design?*

Interviewee: By design, yeh.

Interviewer: *There's nothing in-service that you might review a capability and go "we need to think about this and this..."?*

Interviewee: I don't know how we would do that necessarily. If we have a problem on something else, then we would look across...

Interviewer: *It might spark your...*

Interviewee: yeh, or if we've got some ongoing testing that causes a problem. But generally because we are business focussed as well, we normally find we have a drive to do something which is worthwhile, and looking for something that might not be there is difficult to fund, just on a purely practical basis.

Interviewer: *You've got time and budget restrains so you can't go hunting....*

Interviewee: Although in my role I am expected to try and do this.

Interviewer: *Do you think your output is used effectively by decision makers?*

Interviewee: Yes I would say so, generally.

Interviewer: *Are there any times that you've come at loggerheads?*

Interviewee: Oh yeh.

Interviewer: *Where they think their judgement is correct and you've had to re-educate them perhaps or advise them you're missing the point, this means this or...*

Interviewee: Yes that does happen.

Interviewer: *Are there any common traits that they tend to miss or overlook?*

Interviewee: I don't know about commonly, it doesn't happen that often because it would have to be a pretty big event for that to happen. I mean I was having a discussion this week which was similar to that scenario, where it's more about the balance of determining whether you're expecting something to happen or not....again it's coming back down to the judgement. If you've seen it happen in a couple of cases, are you then expecting that on all the affected fleet, or are you not expecting it because you've only seen it on a couple of cases.

Interviewer: *So is this where the decision makers judgement may be "well we've only seen it in one, why should we ground the rest of the fleet"?*

Interviewee: But how many have you looked at?

Interviewer: *Well, one.*

Interviewee: Therefore you can get in that loop!

Interviewer: *You said it doesn't happen that often?*

Interviewee: No, no because hopefully we don't have too many....I mean these are issues where we would be considering safety issues. It's unusual.

Interviewer: *[Discussion about helping decision makers, motivation for the research, and an anecdote about a personal experience.]*

Interviewee: So in that scenario, the first thing I would do would look at the exposure of those aircraft, so what's their usage been, be that flight hours, cycles, rating, and you stick them in the matrix, plot them on the graph of flight hours and flight cycles, it's what we normally do because our usage is a lot simpler than yours

Interviewer: *discussion about usage metrics*

Interviewee: we would tend to plot all that data in terms of where the aircraft are, show where the affected ones are, try to work out which ones had maintenance actions that should have found it, and you can put them in a bracket to say "it hasn't happened to those for the last x number of flights", so it's putting risks around, try to minimise the risk, understand the risk to the fleet.

Interviewer: *Is that your method, or is that a company/organisation...?*

Interviewee: My method is to draw it, but the way we would tackle it as a strategy would be a company way of doing it.

Interviewer: *So there is a risk management policy...?*

Interviewee: It's not a policy no, so probably it is my method. But I would say I can own that.

Interviewer: *[Discussion about the lack of policy for structure, ambiguity, risk attitudes and capturing information in written form.]*

Interviewee: Sometimes you've got to be careful when discussing these things and we very often find all these non-quantitative discussions of risk are held verbally, formulated verbally and then written down with the people who formulated it verbally, just due to business risk.

Interviewer: *[Discussion agreeing the point from the interviewers experience and the lack of evidence relative to regulated risk threshold levels.]*

Interviewee: The thing I try to do always do, I think this is recommended on our corporate liability course because this all comes back in to liability issues, is to make sure that you have rationale for everything and as long as you can show that you've done things based on your best information, that actually you've done something that might be

wrong, but you've tried and you've done something as best your can, based on all the information you had available at the time.

Interviewer: *[Interview wrap up, summary and thanks.]*

Interviewee code	SIRM-200320191200
Organisation industry	Rail
Organisation category	Regulator/policy maker and operator/end user
Role when handling risk	Advisor

Interviewer: *[Contextual discussion about professional background, and rail industry. Description of organisation function and roles.] We talked about [company name], did you have a risk management framework operating for structures?*

Interviewee: Yes.

Interviewer: *specifically for structures?*

Interviewee: As we mentioned before we started, we are obviously tasked with managing our risks ALARP. How do we evidence that we're managing risks ALARP? The framework they have is, in effect they have a policy they have to meet with a set of standards, then the professional head would set a series of reporting KPIs against which we would report that would be almost a proxy for risk, if you like. So if you're looking at compliance, examination, things like that.

Interviewer: *Quantitative targets?*

Interviewee: Yes, so how many examinations have you hit in a period, compared to a target....but that doesn't really give you risk, it just gives you some KPIs associated with risk...

Interviewer: *It's looking backwards not forwards?*

Interviewee: Quite. So what we then also do, if you go down the next level in terms of managing an asset, you go to the individual component level, each examination looks at the condition of each component, and any defects found are scored against a corporate 5x5 matrix...

Interviewer: *Based on subjective assessment?*

Interviewee: ubjective assessment, I'll come on to that in a minute about how I did some work to try and sort that out because it's not great frankly. So, again that's looking back at what you've got, but they've also got 15 years worth of data now, of what defects occurred, and alongside the defects which are risks, they also collect a condition marking score for bridges...

Interviewer: *I think I've read a paper on this, some work with Sheffield uni?*

Interviewee: Yeah various have been involved. They actually employ a professor of asset management at Nottingham which I'm happy to put you in touch with if needs be....who has lead a lot of the development of using some of our data to do predictive modelling....I co-wrote some papers last few years with some of his team .....so effectively you have all this data you collect at element level and then going through some models you can come up with a predictive model for a future state to try and understand, a bit like you mentioned earlier about when you actually need to examine, when you need to examine, when you need to do the work, when you need to intervene, try to understand at what point you should go and look at things again. So that's all about condition, so you have the backward looking and the forward looking modelling, and when you're looking at setting policy, you use the forward looking modelling to understand across the whole portfolio whether you're dropping condition over time, whether you're holdling the line or whether you're making things better....to try and move towards

some kind of whole life costing structure, which is difficult when you have a bridge that often has no end date. So you know in a component for an aircraft, you have a manufacturer saying it has a certain life, you know you can probably push for a bit, but it would definitely need replacing at some point. With a bridge, we don't know that really because a lot of them are 160/170 years old , we're patching and mending, there's no real need to replace them

Interviewer: *Is there any data anywhere for lifing of bridges?*

Interviewee: Not in terms of life expiry, so that's what we're trying to understand. When the defects occur we do a failure modes effects analysis to understand what happens with that, so we try to do RCM...

Interviewer: *Reliability centred maintenance?*

Interviewee: ...on masonry bridges, but unlike something that's mass produced in a quality controlled environment like a light bulb, when it starts flickering you might know that you've got a day left or whatever it is, and you can fix it. You find that often you can find a lot of defects with masonry bridges, and... no one really knows how long it is, there's so many variables until that actually means something, and often the defect is a positive not a negative. So this wall for example, if you have a crack appear in this wall because some settlement, that's good because it's relaxed all the stresses out in the crack. So getting to the point where you are mature enough across your portfolio to understand when to intervene is really difficult, especially when not only are the structures individually produced, they're in unique environments as well. And I realise that an aircraft is a unique environment, but this is like static, some stuff you may have read that John Andrews is doing, you have a lot of metallic structures next to the sea, lots of saline environment, you have to have a different policy potentially for those than you do for one that's inland on the side of a mountain that's dry. So what they've been trying to do over the last 15 years is collect data to try and bring an acceptable maturity to their policy, so you can drop things down, rather than managing a portfolio of bridges, you've got to then go "I'm going to manage just arches", or "just underline arches with this particular profile on this particular line."

Interviewer: *Is that all to be able to quantify the metrics...?*

Interviewee: To a degree, we had at that time, in terms of investment, we had to justify that we could, we came to an agreement with the regulator about whether you hold the line, whether you manage decline or whether you make investment, but you had to have evidence to justify why you thought that was the case, so it was all about trying to feed the model, try and make things better.

Interviewer: *So in this context of managing the risk, essentially unquantifiable risk, would you class yourself as a decision maker, or an analyst, or an advisor...?*

Interviewee: Personally, an advisor to the professional head because it was his accountability ultimately.

Interviewer: *So he was the guy who signed off, he was the responsible...?*

Interviewee: Yep, but he would fall back on "you've done all the work".

Interviewer: *But he's accountable...so if somebody died, he's the guy...?*

Interviewee: He's the guy who's named on the safety management system.

Interviewer: *As an advisor, did you have any delegated authority, like professional status delegated authority?*

Interviewee: Yes, as I remember it, the safety management system has a single named individual which was my old boss for structures. There was also three principle engineers that worked directly for him, and we had deputy professional engineering responsibility so we could act on his behalf in effect without him requiring direct view of it

Interviewer: *But that was essentially a delegation?*

Interviewee: Yeh, he was happy that you were competent, but it would still be him in court if it was the worst to happen.

Interviewer: *So your decisions that you oversaw or took, did you have to demonstrate this to a professional level?*

Interviewee: In lots of ways yes, but it's not quite a yes/no, binary answer.

Interviewer: *Was it all set internally?*

Interviewee: In agreement with the ORR.

Interviewer: *So they didn't define?*

Interviewee: They didn't define, but they would review and ensure they were content and challenge

Interviewer: *And this was purely because the context that you're operating in, there's very little evidence for life...*

Interviewee: Life expectancy of components. There are some, that's not to say there aren't any. Things like parapets, some parapets, bridge bearings, bridge depth water proofing, things like that that are proprietary products that do have a life expectancy. But most like the world that you've described in the air force, we would never meet the manufacturers required replacement time because we'd go and look at it there'd be nothing after 25 years, so why would you replace it. But at the same time, you have a structure in the other hand that has all kinds of nasty defects that you need to address from a safety perspective, so straight away in terms of priorities, the manufacturer recommended maintenance gets dropped down the list of priorities because you've got limited resource. The other thing I've not mentioned, so they also manage risk in two measures. I spoke of visual condition management. In the bridges space and potentially in the future increasingly in things like culverts, they also look at capability, and what they mean by that is load carrying capacity. So going back to the Victorian times when a lot of the structures were built, the average ages is probably about 125 years, the majority old items, it was built on rules of thumb, luckily the Victorians tended to over-engineer things, particularly structures. So most of the structure is still useful today, but it hasn't got a set weight limit for a modern train, so any structure that was built before 1976 (Which is when it's universally recognised that modern design standards were implemented), any bridge built before 1976 is subject to 18 yearly external inspections, which ranges in three levels. The bottom level being really quick and easy, relatively simple analysis. Level 0 which is ....quite quick and simple, Level 1 is 3D analysis but still relatively simple, and if that still doesn't provide you with the confidence that you need to be believed with the results, if you like, you go to level 2, which is finite element analysis.

Interviewer: *We don't, we go to the OEM*

Interviewee: So *[company name]* have sub-contracted it, but only because they have chosen to outsource it, but that would be a responsibility of *[company name]* to do that. So you have two things then: what's the condition of the asset, and also what's it's capability. And you

have to look at both to be able to understand what the risk is to that asset.

Interviewer: *There's no point in saying you've got a risk, when you're not anywhere near the operating capability...*

Interviewee: Quite, or they're operating in some point at such fine margins, where they say well I've got a risk here, it's below the capability, so what do you do? We just slow trains down for a couple of years, potentially, put a temporary speed restriction on to reduce the dynamic factors, and therefore you're ok again.

Interviewer: *Is that mitigation quantified or is it more subjective?*

Interviewee: It's quantified, it's part of the capability assessment, when you do the capability assessment you look at different line speeds and the different impact of the different line speeds to different scenarios, and then you can make an informed decision on...so the idea is you do the condition bit first, the condition bit feeds the capability assessment, the capability assessment gives you a series of output scenarios, and then you make a decision on the future of that asset.

Interviewer: *Are there uncertainties and assumptions?*

Interviewee: Absolutely.

Interviewer: *In the planned quantitative space?*

Interviewee: So there's interpretation of it, a lot of it comes back to, there's the input would be all bar...so how confident are that the drawings you've got represent what's actually on site bearing in mind a lot of them are hand drawn from 1860. Then there is how subjective is the input capability assessment we mentioned earlier, and *[company name]* have been working on gauge R&R studies to look at how reproducible and repeatable those conditions assessments...

Interviewer: *To see what influence human judgement has on...?*

Interviewee: Yeh, so if you've got the same examiner going in week one and send them back in week four that they'd pick out the same things. Similarly if you send a different examiner in week four they'd pick out different things because of human judgement. Also you need to remember that perhaps for some of these assets particularly if you look at tunnels, if you imagine a mile long tunnel in the dark, it's damp and it's cold and it's the middle of the night and you've got torch light and maybe some lanterns, how confident would you be if you picked up every defect in that tunnel, unlikely right? And that's what they were starting to find it was closed subjectivity.

Interviewer: *And the information feeding their risk model might be garbage or significantly variable?*

Interviewee: Yeh, so they're doing more and more work on that how do they ensure that, critically, how do they bring in more new technology to get rid of the human element.

Interviewer: *Do you think those uncertainties and assumptions in the information that supports the risk assessments, is that communicated to the chief engineers?*

Interviewee: Yes.

Interviewer: *So they're aware of all of them or just the select ones that you tell them?*

Interviewee: So the way it works on an individual asset basis is in the routes, there is a route asset manager for structures, he is the decision maker on a day to day basis as to what to do on an individual asset

level. He or she is fully aware of the limitations of the data that they're getting in, and is also an expert in managing their assets. So they take that into account when they're trying to make decisions on what to do. Then my ex-boss, the professional head, he makes decisions on the portfolio around what's the normal standard for an intervention, or policy for an intervention on an asset. He's also fully aware of the limitations, but he's the person who puts in the additional assurance improvement to try and rectify that limitation in terms of input value, so examples would be the way that they examine a multispan tall masonry viaduct, is at night, as that's only time you can get on the railway.

Interviewer: *Without impacting on operations?*

Interviewee: Yeh, which just isn't tenable for that business, and it's the way that things have been inspected for 100 years. So they go along the top of the viaduct every 5 meters they drop an abseiler down the outside to look with a hammer-tapping survey to look for delamination , voids or whatever we get. Two years ago we did some work with the new drone operator that does inspection, and they flew some of the same viaducts and very quickly found that most varied defects were never found in the last 100 years, long standing defects were never found. Because if you're dropping people every five meters, that persons got a limited field of vision in the dark, so you're not going to see anything. So what they're getting to now in their work is how do you get a point where you send a tactile examination, so you send a drone over the structures and then send someone to the defect, rather than send someone to look at everything because there's obviously implications, and it's dangerous work to a degree, so we're limiting the number of hours on site of the people, and secondly balancing that off with actually how do you make sure they look upon things that are important for a human to make a subjective decision on. So we need to guide them towards a subjective decision, rather than blinding them with mile upon mile of tunnel where there's nothing actually that wrong.

Interviewer: *So you're doing what you can to make subjective judgement more consistent?*

Interviewee: Yep.

Interviewer: *[Discussion comparing aviation and rail contexts.]*

Interviewee: In terms of examination regime. Every structure has an annual visual examination. The standard says something like "every structure will be viewed from a position of safety". Fantastic subjective clause within the standard, what does that actually mean? Well in some cases that could mean from a hillside with binoculars. All that examination is trying to achieve is whether they are major, not structural issues in terms of bridge deformed, more if a handrail has fallen of, is the light adequate, more that kind of space.

Interviewer: *So there's no inference about the integrity of the structure?*

Interviewee: So what they should be doing is checking the defects from the last detailed examination. But often from a position of safety you can't do it. So the examiner might report "I can't see it". So when it comes back to the route asset manager they can see that that defect hasn't been seen for 2 or 3 years, and that's the catch for people to say from a position of safety we can't see this, so the detailed examination might need to be more frequency, perhaps. A detailed examination takes place on average every 5 and a half years, it's on a risk basis. It can be annually and it can be 18 yearly

depending on if it's an inert asset. It's carried-out on all assets again, and in effect it's a tactile examination where we expect to examine everything. But as I said earlier, in some instances where you've got rope access it's difficult to do. Has the route asset manager got confidence that it's been everywhere? It's very difficult to get that confidence. So they're the two main types of examination and how they work together. There are other examinations that the route asset manager has at his disposal as he can request. There's an additional examination, where you've got defects and you're tracking that. Every week, or day, or every hour if it's really urgent you look at it.

Interviewer: *On any of the problems, where you can't go and inspect it, how do you deal with that?*

Interviewee: There would be other mitigating measures you can take, imagine not getting access every night. We've got the option of putting our own condition monitoring on, crack width gauges, tilt monitors, that you can get second by second reporting if you. You can install cameras. But in combination with that, we might reducing speed over the structure, we can look at diverting, reducing to timber line running, operational constraints to manage that risk in the short term.

Interviewer: *If it's an unquantified risk, how do you know that mitigation is going to be effective?*

Interviewee: When you say unquantified what do you mean?

Interviewer: *Something that you've spotted on an inspection, you've found some damage, you can't go back an inspect it.*

Interviewee: If it's that bad, we'd send it to the assessment team to do a capability assessment on it, so you'd get a certain quantified impact on it. Then best practice, not what happens all the time, but what we'd expect, as part of your initial examination regime you'd put a periodicity on it in terms of interval, but in terms of total length. So we expect an intervention at this point, if during this period you see these things happen, we need to do something else, and this is the emergency arrangements if you see this. Thresholds within the monitoring.

Interviewer: *Are there ever any periods where the risk is not fully assessed, and you therefore have to put mitigation on a qualitative basis?*

Interviewee: Yes, let me give you an example. Scour of bridges... *[Discussion of scour, River Crane, in Felton, 2009, RAIB investigation; Lamington Viaduct on the West Coast Mainline following a flood event, costing GBP40mil.]* Scour is not a new phenomenon to structural managers, it is a known risk, it's been known since the dawn of bridges. The problem is, it's pretty difficult to quantify it. *[Company name]* have worked over the last 20 years with various consultants, hydrologists, structural engineers to come up with a risk scoring methodology to understand when you can do something. So the things that influence scour are depth of foundations, direction and speed of water course in very basic terms. So they came up with a methodology to come up with a risk score and after the River Crane incident, it became obvious that *[company name]* weren't as sure in using it as they should be.

Interviewer: *Just paying it lip service?*

Interviewee: I don't think it was that but, it was certainly something that was lower down the priorities, because at a similar time they had a structural collapse due to condition of metallic elements, so there

was a load of stuff going on at the time and it's difficult to maintain visibility of all the risks. But that prompted a load of work to try and understand what the portfolio risk is like. We then in conjunction with experts set a 1 to 20 scoring mechanisms, and we set a line that any that were over 16 had to be mitigated, and what mitigation normally looks like is.

Interviewer: *inaudible*

Interviewee: So in terms of what do you do in the medium term, when you haven't done the mitigation and you know you're carrying a risk, for each of those structures they have emergency plans, so at times of flood they have weather stations and river gauges and in extreme cases water flow rater monitoring to understand at these high risk sites when there's an issue. If they get data to say that flood's happening, they'll send divers in and shut the bridge. So they don't know if anythings happened but to remain safe they'll err on the side of caution, rightly, so until the flood's passed, shut the bridge and inspect the condition of the foundations following the event before we can reopen.

Interviewer: *So it's effectively a conservative approach until you can prove otherwise?*

Interviewee: And that is largely what they do. Obviously you're dealing with hundreds of members of the public over structures that if you've got any concern that they may not be supported, I'd never be comfortable doing that. So they err on caution normally.

Interviewer: *But there's still a qualitative judgement based on the intervention point*

Interviewee: Yes, so the 16 might not be right. So what they're doing now, now they've done another 10 years since that, they're looking at how they can use some of the data they've been collecting. As well as doing that risk assessment, they also have periodic underwater examination of the structure, whether or not it's been flooded or not.

Interviewer: *Just to make sure their' assumptions about erosion are correct?*

Interviewee: Correct, so they're doing some work, or they're planning some work to try and evidence that 16 is appropriate using sensitivity of the model that they've found. Their targets aren't set in stone, if they find evidence that 16 is a wholly inappropriate figure that it's too conservative even, then they'll move it up. The regulator not only pushes to make sure safety is primary, but secondary we are running the safely efficiently. So if we're shutting the railway all the time for a risk model that isn't working, that isn't right either.

Interviewer: *Because the idea of risk is to balance...*

Interviewee: Performance, safety. *[Discussion about analogous aviation perspectives.]* I think you're in a slightly different world. If I found a crack in a structure...unless it was on an element or caused by something I felt was portfolio wide or at least structure-type wide, then I wouldn't go looking elsewhere to a large degree.

Interviewer: *So there's still an element where you as a decision maker or an analyst who works for you will determine an issue is going to be fleet wide, you've still got human judgement in there?*

Interviewee: Absolutely. This wall, if I find a crack in this wall, I'm not going assume that I have the same crack in any of these walls, or in a building up the road because it's happened here and it might be something localised. But an example of something where we would have a portfolio-wide impact is things like management of post-tension concrete. The Italian viaduct collapse, potentially, and

I've not seen the details, this is only my personal opinion, looks like a management of post-tensioning issue. *[Discussion of post-tension concrete construction, it's popularity from 1950s to 1970's and the issues stemming from voids within the grout around the metallic tendons which was intended to protect the tendons]*. So *[company name]* perceived a relatively low risk in that area, compared to people like *[other company]* who have tonnes of 1960's bridges, they might have relatively few, but there's still a risk there. So when the viaduct went down, I used this as an opportunity to raise the issues, so they're not on with a portfolio wide piece of work....

Interviewer: *Post tension structures?*

Interviewee: ...To better understand what the risk looks like. At the minute, it's relatively subjective. You understand where the risk is, you understand how it might manage better than surface defects, but you don't actually know whether you've got a tendon with a critical issue. So what you do is go through a series of deeper and deeper understanding with the asset, you'll categorise and say "the ones from the 1950's are highest risk because they're older, and less time controlled, and we've been through all the history and got some experts in who were there at the time, to understand what the risk factor would be. Prioritise everything, and the top ones, go and do a post-tension special inspection, so that'll be go and do a coring, it's not ideal because you're...

Interviewer: *Interfering?*

Interviewee: ...Yeh, but you core in to various points predetermined on the drawings, to find out whether there is voids or not. There are non-destructive methods but they're not very accurate, then based on the results on that, then you look at well I've got voiding here potentially, what does that mean? How many are around, how many strands of the tendons might I have lost, can I see any I have lost? *[Discussion comparing analogous risk analysis situation in aviation maintenance for context setting.]* Let me give you another example. A place called Stewerton in Scotland, 2009 or 2010. You had three main metallic girders and a ballasted tracked. So the centre main girder had ballast most of the way up it's web. Consequently the visual inspection wouldn't pick up

Interviewer: *Because it was covered?*

Interviewee: It was covered, they wouldn't routinely expose it, because digging away the ballast takes a long time and is a huge operation. And this particular bridge collapsed in 2010 because there was no web, it had corroded through. You're talking major corrosion it's 6mm thick.

Interviewer: *Just galvanic corrosion, or something more insidious? Dissimilar materials?*

Interviewee: No nothing like that, it was painted originally, and it was in a splash zone, so it was wet/dry, and you just had simple corrosion. It's referred to as letterbox corrosion, you can see it just beneath the ballast line and there's just strips of holes missing. If you'd uncovered it, you'd know about it very quickly. So that drove a programme to say where have we got a problem, lets categorise all the structure, let's categorise risk, let's prioritise them and work down the list looking at how bad it is, how often have we looked. So then they have a program of exposing...you can't look at everything immediately because it's operations, so you go for the primary stuff and you go for a sample. So they'll go for two bays across the

length of a structure where they feel that there might be the worst dry area or whatever.

Interviewer: *And that's based on an expert opinion?*

Interviewee: Yes.

Interviewer: *This sounds very much reactive, so you'll get an incident and react to it and you'll see how widely it applies. If you're in a position where you feel this is occurring, is there any proactivity?*

Interviewee: Absolutely, there are things on the risk register ...that haven't happened but we....so metallic fatigue...no one, potentially in the world, has metallic railway bridges as old as the UK. Conveniently everyone in Europe had theirs demolished in the 1940's. So we know that metallic fatigue is a thing, we've got some evidence that it's starting to become a problem but we don't know.

Interviewer: *Low cycle fatigue, or high cycle fatigue?*

Interviewee: What do you mean?

Interviewer: *[Discussion about fatigue drivers.]*

Interviewee: So it's probably low cycle, the issue you have is they're 125 years old...how many cycles has it had? No idea. And again, we were developing a handbook of susceptible details, thinking about Comet - don't have square windows. So set details where you guide an examiner to look for a crack.

Interviewer: *The difference here is we now trust our manufacturers to design out...*

Interviewee: So new bridges are exactly the same.

Interviewer: *But you have old stock so you can't do anything about that, so you just have to take mitigations where possible, and I'm assuming you don't have categorisations of every bridge?*

Interviewee: Yes. They're getting to that point. There will be a data against every structure about what fatigue prone elements they hold, so when you go in to it you can have a look at it and make sure you're ok, and when you do a capability assessment you take into account the fatigue assessment at the same time to understand the remaining life based on what you know.

Interviewer: *Are they two different guys then?*

Interviewee: No they should be the same people.

Interviewer: *We talked about specific examples and cases....but do you have a set process for structural integrity risk assessment, if you can talk generically, or is it all dependent on the circumstances?*

Interviewee: There is a single point of decision making in the process referred as Structural Evaluation. Where, if you imagine, you do loads of research ....and you get one piece of data in, and based on that entire landscape of history and your one piece of new data you make a decision. So in terms of risk, that's the opportunity for the route asset manager to thoroughly understand the risk to that asset and whether they need to make an intervention or not.

Interviewer: *So would you say your risk analysers are heavily dependent on experience and knowledge of the network?*

Interviewee: Yes, they have to be because the stock is so individual, unique, and they know their patches "oh in the summer that bridge has a crack open up, seen that in the last 20 years and nothing happens, it's a stress reliever" or "I know that I can get an extra possession to go and interrogate a particular asset because I know it won't effect the service lines, so we do rely highly on that. To a degree, based

on the type of asset, that's a really good thing, in the past they tried to put more uniform reporting around structures, but that doesn't help it drives perverse behaviours. Because you haven't got that personal touch of "I've been there, and seen it and know exactly what's happening", often as a more senior member of staff you get the wrong sense of comfort.

Interviewer: *Because the paper says it's alright?*

Interviewee: Yeh.

Interviewer: *Is that driven by corporate requirements?*

Interviewee: Potentially

Interviewer: *Anything else that's driving that?*

Interviewee: It's easy ...to get a binary yes/no answer, when in the structures case there is no such thing.

Interviewer: *It's grey*

Interviewee: Yes, it's all grey and it's very much down to a person, in this case my boss *[inaudible]* data management and whether they're comfortable to live with that risk.

Interviewer: *This is where I'm trying to operate in. You have structural experts who know what they're talking about, and they know there's a potential problem here, we believe it's safe.*

Interviewee: And how do you escalate it?

Interviewer: *Well yes, escalate it with the understanding that it's safe if you operate in this space, is that what happens? Do you think decision makers are effective in using structural risk assessment, despite the need or corporate push for yes/no?*

Interviewee: Yes absolutely, the reason for the yes/no, the route asset manager reports to the director...in that space they look after all disciplines. Where I think every multidisciplinary company struggles is comparative risk, so I've got a defect on this wall, and a defect on this bulb, how do I compare the two, where's the common risk currency and they're always striving to get that comparative risk.

Interviewer: *The fact that every risk is unique, do you think that's pushing a difficult agenda to communicate uncertainties to decision makers?*

Interviewee: Yes, potentially but they're doing some good work in that space. They have a tool they've developed over the last 5 years that they're gradually rolling out across different asset types, called the common consequence tool. It's not risk, it's consequence. But it means if I have a failure at this point, this is the potential consequence, and it allows better visibility for a director to understand "this guy might be screaming this is going to fall over, but the consequence is so minimal compared to this" then I can have a discussion between the two about likelihoods.

Interviewer: *So you do try and put a likelihood on those?*

Interviewee: The individual assets will always have a likelihood but it's not part of the common consequence tool yet, they will eventually move to a point where they try and bring that in.

Interviewer: *How accurate do you think the consequence and likelihood is?*

Interviewee: In a lot of the railway the likelihood's pretty accurate, where you have loads of manufacture data saying "this is generally when stuff fails" and you can say "we can push this a bit because we have to"...it's a bit like light bulbs, I know normally this bulb lasts 4 years and in 4 years I've got a 20% chance of failing and that depreciates over time. If you're in the signalling world, that might be the case

because you've got signalling systems that normally last 7 to 10 years.

Interviewer: *Systems versus structure? And this is why I'm focussing on structure because it's so wooly.*

Interviewee: Yeh, so the likelihood bit is a nightmare for structural engineers because a lot of things don't happen instantly, you don't get feedback, and everytime a structural engineer says "we need to something about that" and it doesn't fall down you lose credibility.

Interviewer: *So what do you do about that?*

Interviewee: I think because often the consequence is higher, you can potentially say ...I'd just be open and say "we've got a major defect, my professional opinion is we shouldn't run the trains, the likelihood of failure is difficult to quantify, all I'm saying is my professional opinion is don't do it."

Interviewer: *Do you think decision-makers operate well in that space where you can't tell them a likelihood?*

Interviewee: It's difficult for decision-makers to see the world from your hill, certainly in this new job I'm looking all discipline and it's very hard to compare the two. You have some things that are very subjective like structures, and you have somethings that should be less subjective like items. And so how to develop that common risk model, has got to be something the industry has to focus on. Some of the work I did at *[company name]* was looking at how you evidence ALARP, and in the rail tech there is a way of doing it. The rail safety standards board have a safety risk model, which use a fatality weighted index, that you can use to try and evidence ALARP through gross disproportionality

Interviewer: *Which sounds subjective still?*

Interviewee: So it's reactive, you have to have data, but you can look at how many times an incident has occurred that's killed people or injured people, or minor injuries to people, or a close call in a hierarchy of events, and you can use that data to put a cost of a life. You can use a cost of a life to generate whether you have gross disproportionality.

Interviewer: *So how do you apply that to unique bridges, where your data perhaps is not relevant?*

Interviewee: You need to be careful about where you can and can't apply it, so it's where you have data.

Interviewer: *So those situations where you have unique situations, you can't necessarily apply it?*

Interviewee: Very difficult. Where I applied it was to do with heights of parapets on bridges and the risk of electrocution for people using the bridges. The bridge over here where you walk back over the railway, you'll have 1.8m high parapet, what's the risk to a pedestrian of touching the OLE beneath if you drop the heigh of the parapet, or if you raise the height, so what is the appropriate level in terms of height of parapet on structures for future maintenance. So that's where I was operating. We've got data because we know how many people have been electrocuted, we know how many people have got close, how many people have trespassed and gained access by a bridge parapet, they have the data. So there's two things, is it subjective and is there data.

Interviewer: *But when it comes to the material failure, it's very difficult to do that?*

Interviewee: Very difficult. So you can do it over time, when I spoke about letterbox failures, got data, got one big failure, got numerous close calls when we found it and had to do something quick. So that's the next step in that risk is to say "what's the current cost of our examination regime in this phase, what's the current cost of the performance impact because we need to go in and look at more stuff versus what's the cost of the risk over time.

Interviewer: *That cost of the risk has to have a likelihood attached to it?*

Interviewee: Yes, but we use the data to evidence the likelihood, how many close calls have we had, how many times have we had partial failures.

Interviewer: *I think that's my uncomfortableness that sometimes structures management practitioners are using numbers to be able to justify courses of action when perhaps the likelihoods might not be in a similar space, and you can't say how accurate you are, there might be a level of conservatism built into the structure.... it's almost arbitrary.*

Interviewee: Yeh. One of my colleagues talks regularly about normalised deviance in the railway. I started my career on the highways. When was the last time you drove down a motorway and saw trees growing out of bridges on the higway? It's never happened because the normal approach to risk is you don't do that, that's not a good thing....preventative. How often do you see it on the railways? You'll see buddleias growing out of the structures, because if you go ask the people in the room, that's the level of risk people operate at, that's not written anywhere. It's a "it's the way we do things around here" type approach, and that used to really worry me.

Interviewer: *And it worries me because it's a natural human instinct.*

Interviewee: We're ok because the person next to me says we're ok

Interviewer: *Absolutely and that's one of the things I'm taking into account because there's so much human subjectivity in structures, that normalisation of deviance is inherent to the way that we do business.*

Interviewee: So the examiners are artisans largely at *[company name]*. They're not engineers, they're eyes and ears. I'm not sure that's the right thing to do, in the highways sector that's not the case, they're chartered engineers. They not only understand the defects, they understand the ramifications of the defect. So you have a crack there, that means you've got high stress, bad.

Interviewer: *Rather than someone, going "we've got that, we've not seen that fail before".*

Interviewee: Not even that. Our examiners will just go "crack", write it down. I'm doing them a disservice, but that's the worst case scenario.

Interviewer: *Do you think there's quite a lot that doesn't get reported that would be useful to be reported?*

Interviewee: Yep, loads of contextual stuff so some of the work I was doing was how do we put 3D modelling in VR, so we back from an exam, you can sit back and go "oh it's there".

Interviewer: *But you're duplicating what the surveyor has done?*

Interviewee: They're not a surveyor that's the problem, they're an artisan. They're eyes and ears on the ground. It comes back to an engineer, but the engineer has to score on a 5x5 matrix, likelihood and consequence, and the likelihood of when it's going to fail or degrade, is all subjective.....

Interviewer: *Unknown unknowns, do you do any activity that formally tries to capture?*

Interviewee: We do horizon scanning with other big asset owners, have they got risks on their portfolio that haven't brought themselves to bare within our portfolio.

Interviewer: *You mentioned FMEA as well?*

Interviewee: We do FMEA.

Interviewer: *But that's reactive by the sounds of things?*

Interviewee: It is, but it's also very difficult to understand, you have to do it on an individual asset basis because they're unique, there just isn't the time or the capacity to do that on an individual asset, so you end up doing it at asset route level, and then it becomes less useful potentially.

Interviewer: *Do you have confidence in the risk assessments that you're given to decide upon?*

Interviewee: Provided they've been through the process, and they have a three step approval, yes.

Interviewer: *Three step approval meaning?*

Interviewee: They have an examiner that's been working on site, they have an engineer who's done the risk assessment of the defects coming back, and they have a senior engineer overseeing the process. That's not saying it's without flaws, I used to be really worried about picking up all the defects on site, which is why they're spending so much money trying to investigate new technology to take away subjectivity. I'm also worried that in the intermediate step, that the engineer doing the review, at the moment they look at everything. If you imagine you get a piece of paper and it says everything's ok, you get another piece of paper and it ok, imagine 20 like that, the 21st one you're expecting it to be ok, it's not and you miss it. That's what I'm worried about. I think the process is right provided the human factors are accounted for, and that's what *[company name]* are now working on trying to understand more decision making and human subjectivity.

Interviewer: *Would you say it's risk informed or risk based? Risk informed is you're using risk, using the output to support a wider argument. Risk based is the "risk assessment says this, we're doing this."*

Interviewee: The second.

Interviewer: *Do you have assurance activity that oversees the risk assessment process, specifically?*

Interviewee: Yes, the supply chain go and do assurance activity on examinations, a certain number of end product checks, then the route will do a sample check, then the centre will do sample checks in a small number to verify it's all working.

Interviewer: *So you have internal and external audit going on to cross check?*

Interviewee: Yep, and then also cross-route, so one route will go and sample check another route's, peer to peer.

Interviewer: *To clarify, routes?*

Interviewee: Operating areas, like Anglia and LNE swap people for a couple of days.

Interviewer: *Lastly, considering all the structural integrity risk management decision making, are there any commonalities between the uncertainties and assumptions that decision makers understand? You've talked about it broadly, but is there anything specifically*

they don't get like they don't understand the human subjective element or they don't understand the crack propagation, or they don't understand corrosion?

Interviewee: Yeh, it's always very difficult to escalate to senior management what you're concerned about, because often it's very techy and engineers like talking about techy stuff, but senior managers don't. So trying to articulate in an elevator speak what the issues and what the ramifications are to them is always tricky, and what risks they are carrying. Hence the need to get to a point where you have commonality of risk tools

Interviewer: *And the need to quantify in numbers they understand. Would you say decision makers are generalists?*

Interviewee: Yes

Interviewer: *Have there ever been structures people in those positions?*

Interviewee: Yeh, but it's not because they were structures people, it's just they were good managers

Interviewer: *Do they go poacher turned game keeper?*

Interviewee: Yes.

Interviewer: *...I'll stop it there.*

Main survey interview transcripts

Interviewee code	101220190940 (ID 58)
Organisation industry	Aviation.
Organisation category	Operator/end user
Role when handling risk	Risk decision-maker

Interviewer: *...are you happy for me to record this interview?*

Interviewee: Yes

Interviewer: *Thank you very much. So the consent for that you signed, ill just run through what the consent is. [Consent discussion]. So do you consent to your responses being used on that basis?*

Interviewee: Yes. I do.

Question 1

Interviewer: *Mega, thank you very much. The first section is quite quick, it's context setting. So what industry is your organisation? Aviation? Probably the most accurate?*

Interviewee: Yes, probably.

Question 2

Interviewer: *What category describes your organisation? Would you say you're Operator/end user, Regulator/policy maker, Designer/equipment manufacturer, or other?*

Interviewee: So at the moment, other because I'm in the training environment. But do you want me to answer as when I was in the PT.

Interviewer: *So from your experience when you were doing structural integrity risk stuff.*

Interviewee: Operator/end user, I'd say, that kind of end area.

Interviewer: *I should have said that at the start, so while I appreciate the job you're in at the moment is not directly using structural integrity risk management...*

Interviewee: I'll refer back to previous.

Interviewer: *I mean that was 2 years ago or there abouts, so it's relatively recent.*

Interviewee: Ish...May 18 I left.

Question 3

Interviewer: *What was your role within the organisation?*

Interviewee: So I was in the *[aircraft project team]* and I was one of the *[airworthiness delegation]* holders making decisions for airworthiness on structural integrity.

Question 4

Interviewer: *In your experience, do organisational risk management policies specifically consider structural integrity risk.*

Interviewee: They usually refer to it, however the detail of how to analyse in full depth the structural integrity risk is usually either work in progress or something that is relatively vague, so whilst it will give you guidelines as to what is a remote risk or whatever, it then doesn't give you a way of actually assigning any probability to that, it does tend to leave that area woolly, whereas with things that tend to have failure rates, it almost seems to be focussed around those at the moment.

Question 5

Interviewer: *With the context of equipment structural integrity, which of these best describes your role or the role that you performed when handling risk. Were you a Risk analyst, a Risk decision-maker, a Commentator, or an Advisor?*

Interviewee: So decision-maker.

Question 6

Interviewer: *So those sorts of answers help me be able to group or identify from the next three sections how you've responded and compared to other people within those cohorts. That's why those context questions are asked. So the next section addresses the analysis aspect of structural integrity risk. Describe the current process for conducting risk analysis for in-service structural integrity issues, from identification of an issue to when you will have considered it managed.*

Interviewee: Ok. So it changed whilst I was in the PT slightly. But from when I was leaving what we would do is we'd get an issue in, and then we would send usually off to the design organisation, ask for a bit of advice, they would give us a preliminary sort of gut feeling, quite often informal over the phone and from that we'd probably make a first decision as to how quickly or otherwise we needed to progress with the risk in hand. They'd then come back later down the line with a formal response in the mean time we may have already written a temporary risk assessment that sort of says the overall risk is unknown but here's our first kind of gut reaction of what we think we should be doing in order to manage it. We'd wait for further analysis from the design organisation that would come back to us, that could either be something relatively small and take a matter of days or we could be talking about going off to work with the other nations and we're talking months or even years to get a PDT *[post design task]* done so that you can come out with a bit of full analysis. Ongoing through all that you'd be updating your risk assessment so we had the *[hazrd management working group]*, which you'll loving remember, which we prepared briefs for, and we'd keep updating risk assessments, we'd assign a probability to the likelihood of that event occurring that tended to be based on a gut reaction or from us as a team we'd discuss it with the *[other company name]* structural advisor, with *[other company name]* but they weren't prepared to assign a probability to it without data behind it. We would then assign a probability to it within the PT based on a set of criteria that we predetermined.

Interviewer: *So two things from that: you said that if you had the long period between getting like a PDT out, would your initial gut reaction that you were then updating presumably your information in, that would still be the mainstay of your risk assessment, so the gut feeling that you had that generated the likelihood, how did you come up with the number for that?*

Interviewee: So probably shouldn't use the word gut feeling should I, I probably should have said engineering judgement.

Interviewer: *It's still judgement, but either way how did you generate that number?*

Interviewee: So what we did, we came up with a bit of a template of risks, so we had the design safety target, which I think off the top of my head was $1x10^{-6}$ which we kind of said if there was no risks on the platform because it was an old platform, there was absolutely

nothing wrong with it, you had a brand new jet that had come straight off the line, there'd still be a risk associated, a structural integrity risk that was kind of unknown, and that would be at the level fo $1x10^{-6}$. On top of that, we'd then have the other risks that were known on the platform that we'd then add in. So we took that design safety target as being kind of (the baseline) the baseline, and then we added into that other numbers that tended to be somewhere in the region of 10^{-7}, 10^{-8}, occasionally 10^{-9} if we really weren't that worried about it, that all added into that overall risk, so that we could come back with a number that the duty holders could use and the *[senior aircraft airworthiness authority]* could use to inform their platform risk.

Interviewer: *So the numbers that you used were generally, would you be happy calling them approximations?*

Interviewee: Yeh they were definitely approximations.

Interviewer: *Comparative approximations?*

Interviewee: Yeh they tended to be similar levels of risk, and other non-structural risks on the platform, so where we had hazard management guide, do you remember what it was called? The *[organisational safety management plan]*, there we go, so we had that and that said in there where things were remote, and whatever the ones below, can't remember, that's bad, yeh we took those bands and basically picked a number between the highest end of each of those spectrums and the lowest end, and used that to base our kind of likelihood of it happening because we realised that previously we were saying that the more the aircraft flew, the safer it got because you'd just say "oh it's never happened, so the likelihood is 1/the number of flying hours", which isn't right because that just implies that you're just getting safer and safer the more that you fly, but yeh, we were just taking an approximation of those.

Question 7

Interviewer: *Ok we'll probably come into that a bit later but I wanted to dig into it a little because you'd mentioned it. Ok so the next question is, there's a few of these throughout, which is between, it's called a likert scale, between strongly agree and strongly disagree, going strongly disagree, disagree, slightly disagree, slightly agree, agree, strongly agree, but I'll remind you of those as we come to them. There's 6 but you have to choose whether you agree or disagree, it's just what weighting, you can't go down the middle. So how far do you agree with the statement: In my organisation, structural integrity risk analysis is an activity that is valued?*

Interviewee: Definitely agree, what is it strongly...

Interviewer: *Strongly, slightly or just agree.*

Interviewee: I'd probably strongly agree that it's valued.

Question 7a

Interviewer: *Why do you agree that it's valued?*

Interviewee: Because it's acknowledged that they're the most significant risks recorded on the platform, so they needed to be given the time by the duty holders and the *[senior aircraft airworthiness authority]* to look at them in detail, so yeh the risk was valued in terms of...I take it by valued you do mean almost appreciated-type valued, rather than valued as in given a number...

Question 8

Interviewer: *Yes valued in terms of appreciation. Ok. Within your organisation, are structural integrity risks that arise in-service described numerically or non-numerically/verbally/descriptively?*

Interviewee: They're both. We're forced to describe them numerically but realistically they are described verbally throughout. It's only when we put them into the loss model that the number really means anything.

Question 8a

Interviewer: *And how would you rate the accuracy of the quantitative risk analysis?*

Interviewee: As in the number that we assign to it?

Interviewer: *So it's between very poor, poor, good, very good.*

Interviewee: Poor.

Interviewer: *Why do you think that's the case?*

Interviewee: The only reason I'd say poor rather than not very poor is because it is, we were basing it on the experience and the judgement of some very experienced people, so people who've been in the industry and around that specific platform for 30/40 years, so the combined experience that was looking at the risk and assigning it a probability was very high, however it was still an engineering judgement rather than it being based on any analytical assessment.

Question 8c

Interviewer: *So another disagree/agree question. In my opinion, a quantitative (numerical) value is necessary for the assessment and management of in-service structural integrity risk?*

Interviewee: It is in the current way that we manage risks...all the other risks across the platform.

Interviewer: *So you need it because the rest of the system is managed numerically, is that what you're saying?*

Interviewee: Yeh.

Interviewer: *So that's somewhere on the agree side?*

Interviewee: I'd probably say agree, but not strongly. Just agree.

Question 8d

Interviewer: *So, In my opinion, a qualitative (verbal/descriptive) interpretation is necessary for the assessment and management of in-service structural integrity risk?*

Interviewee: Strongly agree, because that gives the context behind.

Question 8e

Interviewer: *Do you have any comments or anything to...ah you've already mentioned some of it, is there anything you want to say that will expand on why you've said those?*

Interviewee: No, I don't think so.

Question 9

Interviewer: *When handling in-service structural integrity risks, are you required to assess structural integrity risk relative to a threshold level?*

Interviewee: Er...yes, in that we have the design safety target and then we also had the number we couldn't go above, in terms of the overall risk assessment of the platform.

Question 9c

Interviewer: *Was that regulator set, or statutory, or an internal organisation, or personal limit?*

Interviewee: I think it was regulatory, I think it was an *[aviation regulator]* limit, or it might have even been an EASA limit that was then read-across to the *[aviation regulator]* that we were then working within. And then because the platform wasn't designed within normal tolerances as in the way that we now design platforms, there was a waiver, a relaxation of that number to acknowledge that actually the target that they were aiming for when they were designing the platform was just a higher than the current day target.

Question 9a

Interviewer: *The nature of that threshold level, was it a loss rate per flying hour?*

Interviewee: Yeh.

Interviewer: *was there anything else that the threshold required?*

Interviewee: I think that was it from what I can remember.

Question 9b

Interviewer: *In operational situations which are information poor, time critical, or resource constrained, or any other pressure that you might have had, I believe a threshold level is appropriate.*

Interviewee: I think I disagree but not strongly. What's the middle one? (Disagree. Slightly disagree, disagree, strongly disagree, that's the order). I think I disagree, because, yeh, the reason behind that is because actually as soon as you put a numerical threshold on it then it almost can blind people to actually analysing the situation and they just go "it's alright it's below the number, crack on" whereas actually certain situations actually you might not want to take as much risk, or vice versa, you might want to take more but if it's documented with a threshold I don't think that gives the true interpretation of the risk versus benefit.

Question 9d

Interviewer: *What processes do you use in support of SI risk analysis or decision making? So I've got a big list, so if I work through them and if you just say yes or no. So, what if analysis? As in you do a "what if this happened...this"*

Interviewee: Yeh I think you automatically do that as part of your assessment don't you.

Interviewer: *Checklists?*

Interviewee: Not really no.

Interviewer: *Safety audits?*

Interviewee: Yes.

Interviewer: *Task audits?*

Interviewee: Yes, well we use them as part of the PTs quality system but we wouldn't do a task audit on a specific structural integrity issue.

Interviewer: *But you would do a safety audit?*

Interviewee: Yes, I guess so, yes, we'll go with yes.

Interviewer: *Bow tie analysis?*

Interviewee: Yes definitely.

Interviewer: *Proportional Risk Assessment Technique. It's a specific thing.*

Interviewee: If it's a specific thing, then no.

Interviewer: *Risk matrices?*

Interviewee: As in like the probability vs ....yes.
Interviewer: *Quantitative Risk Assessment?*
Interviewee: Yes.
Interviewer: *Probability distributions?*
Interviewee: Yes. Ish. I'm trying to think of when we actually used them for anything.
Interviewer: *So that's where you have a...you expect it to follow a set distribution, a set pattern and you're using that as part of your assessment.*
Interviewee: In fact yes we did, when we had stuff back from PDTs, not in an initial sense.
Interviewer: *Event data models, historical data, frequency of failure, that kind of stuff?*
Interviewee: Yes, but not often. I mean yes, we used to hunt for historical data but yeh.
Interviewer: *Human error analysis technique?*
Interviewee: Is that a specific technique? No.
Interviewer: *Fault tree analysis?*
Interviewee: Yes.
Interviewer: *Event tree analysis?*
Interviewee: Yes.
Interviewer: *Risk based maintenance?*
Interviewee: Yes.
Interviewer: *Epistemic models?*
Interviewee: I don't know what that is, probably not.
Interviewer: *Bayesian Networks*
Interviewee: Again, I don't...no.
Interviewer: *Monte Carlo Simulation Techniques*
Interviewee: Nope.
Interviewer: *Bayesian Analysis?*
Interviewee: No.
Interviewer: *Probabilistic distributions for key variables? So in a structural integrity sense that might be your, it might be a probability distribution for your crack propagation rate.*
Interviewee: Ok, then yeh.
Interviewer: *Bespoke numerical models for a given situation?*
Interviewee: Yes.
Interviewer: *Neural Networks.*
Interviewee: As in brain?
Interviewer: *As in machine learning-type approaches.*
Interviewee: Probably not, no.

Question 9d1

Interviewer: *Do you believe these processes are appropriate for structural integrity risk analysis in your organisation?*
Interviewee: As in all of them, or just the ones I've mentioned?
Interviewer: *Just the ones you've mentioned.*
Interviewee: Yes.

Question 9d1a

Interviewer: *Why do you believe that's the case?*

Interviewee: Because I think you need some kind of structure to focus your risk analysis around, because without some or all of those techniques you're kind of just putting a finger in the dark and going "it's about there".

Interviewer: *Did you use all of those, or were you using someone elses, like the OEM had used those methods?*

Interviewee: Yeh, so you'd get information from... (saying that they'd followed this process?) Yes.

Question 10

Interviewer: *Is your organisation's risk analysis methodology proactive, enabling anticipation of risk from in-use changes or modifications to usage/configuration, or is it reactive, where you're responding to issues that occur?*

Interviewee: I'd say on the most part it's reactive because you're responding to issues that occur. The only time that I would say it's proactive is if you're doing in-service inspections, looking for issues because you think you might find them in a certain area or when you're doing teardown or something along those lines. But the majority of the structural integrity stuff that we dealt with was reactive.

Question 10a

Interviewer: *What methods did you use to identify risks?*

Interviewee: So there was major, minor, minor star servicings. Feedback from front line Squadrons from when they found issues. Discussions with the other nations. And then teardown and MAFT analysis.

Question 11

Interviewer: *Another agree/disagree one. The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'likelihood' of occurrence.*

Interviewee: I'd say disagree because for the majority of cases, it's a clear likelihood in that we've said "it's this" but it's not an accurate likelihood.

Question 12

Interviewer: *The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'consequence' of occurrence.*

Interviewee: I'd say I agree because we're doing the event analysis to say "if this happens, x, y, z will happen". I wouldn't say strongly agree because there's things you don't foresee in those kind of situations.

Question 13

Interviewer: *More often than not, there are uncertainties within my analyses of structural integrity risks. Uncertainties here are classed as information or a value that is not known accurately.*

Interviewee: Strongly agree.

Question 13a

Interviewer: *Are uncertainties communicated in risk analyses regarding operational structural integrity?*

Interviewee: Strongly agree

Interviewer: *That's a yes/no, sorry.*

Interviewee: Oh, Yes!

Question 14

Interviewer: *More often than not, there are assumptions within my analyses of structural integrity risks.*

Interviewee: Is that a yes/no?

Interviewer: *It's a strongly disagree to strongly agree...I need to get better at this interviewing!....an assumption is an activity or value that is presumed to be true, for the purposes of analysis.*

Interviewee: I'd strongly agree because again we were using probabilities that were all based around that initial design safety target.

Question 14a

Interviewer: *And are those assumptions communicated in risk analyses regarding operational structural integrity? Yes or no.*

Interviewee: Yes.

Question 15

Interviewer: *More often than not, I handle structural integrity risks conservatively. Conservatism here is cautionary thinking, over-estimation of the risk resulting in additional measures or values added to provide a safety factor given uncertain quantitative outcomes.*

Interviewee: Yes, you have to don't you other you're being a little bit reckless.

Interviewer: *that's an agree/disagree*

Interviewee: I do agree, yep, because there are situations where you wouldn't be conservative but they'd be the times when it's, when it's saying there's an operational necessity that we need to be doing something that we means you're not being conservative.

Question 15a

Interviewer: *And yes or no, you communicate those when you're being conservative?*

Interviewee: Yes.

Question 16

Interviewer: *To what extent do you believe human subjective judgement (from expert opinion or belief) influences the analysis of in-service structural integrity risks? This is a slightly different scale: Not at all considered / To a very little extent / To some extent / To a great extent / To a very great extent.*

Interviewee: Can you read the question again.

Interviewer: *To what extent do you believe human subjective judgement (from expert opinion or belief) influences the analysis of in-service structural integrity risks? Human subjective judgement here is a judgement influenced by human opinions, feelings or tastes.*

Interviewee: Er....not the highest one, but the second one down.

Interviewer: *To a great extent?*

Interviewee: What's the one below that?

Interviewer: *To some extent.*

Interviewee: That one.

Question 16a

Interviewer: *Are human subjective judgements communicated in risk analyses regarding operational structural integrity? Yes or no.*

Interviewee: Er.....it's almost to some degree, I think it almost depends who's writing it. And some people's almost unconscious bias won't be necessarily conveyed, so someone might be more conservative than someone else, and others have known the name at the bottom

of the piece of paper once you've got to know them, there will be nothing in there to necessarily determine the level of conservatism that is being applied. Only the bigger reasoning behind the decision but not necessarily anything to quantify why the decision (taken in that manner) yeh.

Interviewer: *So that sounds more like a no?*

Interviewee: Yeh, er...in some cases it was but yeh, so for example you might say "I want to inspect that every 25 hours", somebody else might say "I'm going to inspect that every 50 hours" and there'd be nothing in there to differentiate...you could have exactly the same written justification for those decisions and the end result could be a different inspection frequency.

Question 17

Interviewer: *So I've put no for that but I've also put "caveated", but this is the beauty of doing an interview is that this the level of detail I can get from it, whereas in a survey you'd have had to chosen one of the two. So another, disagree/agree... I am confident in the structural integrity risk analysis method applied for handling operational risks. Operational in that sense is when you're constrained and up against resources, potentially with information..deprived of information.*

Interviewee: So I think in it's current form it does the job, so yes, but however, caveat it again with there's quite a lot of areas for improvement because going back to what we were saying before, you could have something that is assessed by one person as being absolutely fine and by another person as absolutely not, and the justifications for those decisions could be the exact same justification.

Interviewer: *So shall I put agree for that, or slightly agree?*

Interviewee: Yeh, slightly agree one. I think it's one where it is fit for purpose because it allows the person taking the risk to make that decision, but there's room for improvement in terms of providing guidance rather than policy.

Question 18

Interviewer: *Do you have any other comments on that last section?*

Interviewee: I don't think so.

Question 19

Interviewer: *So the next section's on, so that's two sections down, two to go, we're about half way...it should pickup from now. What information is essential in order to make a decision upon an in-service structural integrity risk?*

Interviewee: Have you got a list? Or just want a spiel?

Interviewer: *No this is open, so if you had a crack on here, then what would you consider as necessary to make that decision?*

Interviewee: So on a fleet perspective you probably want to know the average usage previously and what it's going to be doing going forwards, the operational environment that it's working in, you obviously want to know the location, dimensions, as much information as possible about the specific fault, any other associated defects in that structure, which is a bit of information that if you're in a project team you don't always get, you just get information about that very specialised "if you've found this crack here", you don't often get "and here's the history of the leg" or whatever it is that you're looking at, yeh, what the plan for the aircraft going forward. And then also, how repairable in-situ is it, is it something that could easily be

replaced and you just take the risk out altogether, what is the operational impact of doing whatever maintenance, or rectification should I say to remove that risk altogether. That's probably about it.

Question 20

Interviewer: *Describe the decision-making process regarding in-service structural integrity risks.*

Interviewee: So it's fairly, almost all the things that I've previously said, take those into consideration. If you can get rid of the risk entirely without having an operational impact, that's obviously the first tick vg (very good) lets do that. If you can't then it's a decision between the assessed severity of the risk versus the operational impact of not flying that aircraft whilst you fix it. And realistically that can vary from being a very easy decision to being a really difficult one that you need to do a lot of digging to get the information for.

Interviewer: *So the difficulty comes from having to justify why you would want to continue flying with that damage.*

Interviewee: Yeh. Yeh. So one of the, I guess an example of that is when we had that, I don't know if you were still in the PT actually, when we had the rust under the mod plate on the wing and we had the aircraft that were flying back from *[operational location]*, and there was an aircraft where we saw on the outside of the plate that there was corrosion, but we didn't know what was underneath it, and it's then a decision of what's the cost, yeh that's another thing, the practical cost, maybe not in terms of money, but in terms of man hours and impact to the ability to look after the rest of the fleet so we had a decision as to whether to send our NDT teams out to *[operational location]* or to keep them where there was lots of operational aircraft flying, in order to fly this aircraft home and check it was safe, so there's also that kind of thing to weigh up as well.

Question 21

Interviewer: *Do you ever use judgement when making or supporting a decision regarding in-service structural integrity risks? Do you use judgement.*

Interviewee: Yeh, I think every time it's engineering judgement and it's based on your previous experience and yep.

Question 21a

Interviewer: *So I've got a list. In which types of situation do you use judgement? When there's a lack of available expertise?*

Interviewee: Yes.

Interviewer: *When time constrained?*

Interviewee: Yep.

Interviewer: *When I disagree with advice?*

Interviewee: Yes although caveat that with if it's advice from the design organisation and you disagree with it, then that would be above the level that I would have been allowed to. However on the flip side of that, if I wanted to be more conservative, then I can do that.

Interviewer: *When there's insufficient resource to support a decision choice that would other be available? Resource meaning spares, information...*

Interviewee: Yep.

Interviewer: *When there's insufficient information available to support an objective analysis?*

Interviewee: Yes, that kind of links into the time pressure one doesnt it, because you wouldn't do that unless you've got a significant time pressure.

Interviewer: *When I believe a particular decision is good enough?*

Interviewee: That makes it sound gash, but yeh.

Interviewer: *There is actually academic work, it's called satisficing, it's the posh word for good enough! When I believe that technical information is conservative?*

Interviewee: Yep.

Interviewer: *any others?*

Interviewee: I think that pretty much covers it. I think in pretty much every situation you're using some form of your own judgement even if you've got a huge amount of technical evidence to support it, because unless you are just replacing a component and it's an easy, you know, you've got a million spares on the shelf, you've got all the time and the resource to do it, you're always having to apply some form of judgement as to when to get that done, even if the judgement is "I've reviewed the advice and I agree with it".

Question 21b

Interviewer: *Is the judgement that you use holistic, which means using your gut, intuition or instinct, or do you decompose the judgement upon the situation into constituent parts.*

Interviewee: A bit of both, but both's not an option is it? Because you use an element of both. You have to break it down, but for each bit you've broken down, you probably then still apply your own kind of gut reaction, holistic.

Question 21b1

Interviewer: *Are the judgements you make numerical or non-numerical/verbal?*

Interviewee: Can I have another both? It depends on the situation.

Question 21b2

Interviewer: *if you decompose the situation, how would you go about decomposing it.*

Interviewee: Probably very similar to the "what information do you need" question, so I'd be looking at the perceived risk or doing nothing. I'd be looking at the ease of rectification, time constraints, operational factor, basically you break it down into the, what if you do nothing, versus all the reasons why you need to do something, and kind of weigh it up against each other. Does that make sense?

Interviewer: *Yep. Do you have any weightings? For those options that you do, do you have like an equal weighting?*

Interviewee: I think that goes back to the question on is it holistic or....i think once you've broken it down, you look at them, some of them may be weighted, others definitely won't be, so it's very difficult to give a weighting to the operational impact of not doing something, it's very difficult to give a weighting to the amount of time and resource that that's going to take in relative terms, i.e. if you've got somebody that's doing a 3 hour maintenance job on whatever it is, they're obviously not doing rectification jobs, they're obviously not doing 3 hours on another aircraft. So all those things, I don't think you can give a quantified number to, so I think you're going back to your military judgement, your engineering judgement, and almost assigning non-numerical values to that to weigh it in your mind.

Interviewer: *Would it be fair to say, and I'm calling on my own experiences here, whenever you looked at a course of action, that you kind of had a good feeling about some courses of action, and a bad feeling, you had some levels in between as well as to which ones you'd prefer, but you wouldn't necessarily decompose that into "well I prefer that because of this particular value" because you perhaps might have felt that it was silly to prioritize "well I prefer that one because it's a 3 hour [inaudible]".*

Interviewee: Definitely, once you're then looking at each individual problem, there's not a specific "I'm prioritizing this one because" it could be for any number of reasons and there's definitely a, yeh, a judgement call rather than a "ah the number says it's this, we'll look at this one".

Question 21c

Interviewer: *So kind of related to that narrative, what do you find difficult when using judgement to support decisions regarding in-service structural integrity risks?*

Interviewee: I think with judgement, there's obviously, you're calling on your own experience, you're calling on the experience of others, and particularly in the way the military move people round, you're always going to have a level of inexperience compared to design organisations and what have you, in some ways that good because that makes you look at things potentially more critically than if you've seen it for the last ten years "ah we've since this for the last 10 years it's always fine" but because you haven't got that "we've seen this loads of times and it's always fine" you often have that questioning of a decision. So you can make a perfectly sound decision but you're always going to go kind of back to it and think was that actually the right call. So yeh, I think that's the main issue with it.

Question 21d

Interviewer: *How could you be supported better, in that sense, when using judgement to support decisions regarding structural integrity risk?*

Interviewee: I honestly don't know, because actually you are well supportrewd in that you can go back to those experienced people, like the decision orgnaisations, your independent structural advisor, the people who've been in your project team for longer, your chiefs flight sergeant and civil servants, so you've got a good level of support there, you've got your peers in the team.

Interviewer: *That's more resource that's available to help you make future decisions. If you're stuck with the decision...*

Interviewee: Yeh, you're just talking to those people aren't you, but ultimately at the end of the day once you've tapped out all that information and you've...you've exhausted...you've just got to make a decision and I guess the only thing that you can do if you're not sure is to go conservative. But again, operational situations might make that more difficult than others. But yeh, I genuinely don't know what you could do to improve that for the decision-makers.

Interviewer: *Send someone on a PhD....*

Interviewee: Indeed! That's what you could do! Send someone on a PhD...put the words in my mouth!

Question 21e

Interviewer: *I'm justifying my existence! Have you received training in applying judgement to support decisions regarding structural integrity risk?*

Interviewee: Yes, in that you do your package that leads you up to get your *[airworthiness delegation]* and you go on the *[aircraft structural integrity course]*, and the *[airworthiness course]*, and then you get on the job training from people who are also in the team around you. Are you sat down and said “lets go through some examples” and talked through decisions that were previously made.....then no. But overall I think yes you are given training, could that training be improved? Potentially.

Question 22

Interviewer: *In operational situations, decisions regarding a structural integrity risk are clearly justified? An Operational situation is a situation that is time critical, as well as information poor and/or resource constrained.*

Interviewee: Erm, I would say that they are usually strongly justified. Probably not strongly agree because there are occasions where they are made, not by me obviously, but by people in pressurised situations where they’re making a decision very very quickly, but actually the majority of them are, even if they’re time critical and information poor, you’re still justifying your decision based on the facts of the information that you known. And I think pretty much every person who is not wanting to go to jail, would justify their decision. Whether they would have the correct information to be able to make the right decision is another matter.

Question 22a

Interviewer: *Is there a mechanism for ensuring that structural integrity risk analyses are justified?*

Interviewee: Yes, in that there’s the *[organisational safety management plan]* that gives you the structure for how we should be assessing risk in general across the platform.

Interviewer: *That’s in general across the platform, what about specifically for structural integrity risks? You said earlier about the degree of detail for structure is not necessarily that clear....*

Interviewee: I think there is a structure in place but again it could be improved because the structure is a structure that is almost designed for, not for structural occurrences, if that makes sense. So you can manipulate the information that you’ve got to go in to the structure that is given...

Interviewer: *The risk management structure, as opposed to the aircraft structure.*

Interviewee: Yes sorry. Yeh so you can manipulate your structural information and your structural risk to fit into the risk management system, let’s put it like that, but the risk management system isn’t designed for structural issues.

Interviewer: *So not designed meaning not appropriate?*

Interviewee: Erm, I wouldn’t say not appropriate, it just could be improved, which goes back to our assigning probability.

Question 23

Interviewer: *Are you, personally, more comfortable with numerical or non-numerical/verbal/descriptive representations of risky decisions regarding in-service structural integrity?*

Interviewee: I think I’m more content with non-numerical, but I think that’s only because I’ve been in a situation where I haven’t really had the evidence for the numerical numbers that have been applied. Evidence is the wrong word....

Interviewer: *Information, or analysis?*

Interviewee: Yeh, we have assigned numerical probabilities to things, and numerical values to things, but actually the verbal justification that supports that number was far more important than the number itself, because the number itself was a judgement based on the verbal rather than the other way round. So I think because of that and working in that environment for 2 and a half years, I'm much more comfortable with the verbal descriptions of things rather than the numerical.

Question 24

Interviewer: *Within your organisation, are decisions upon structural integrity risk assessments "risk-informed" or "risk-based"? Now Risk-informed is Risk assessment results support decision making but are not the only consideration. Risk-based is Risk assessment is the sole basis for decision making.*

Interviewee: I think mostly they're risk-informed.

Question 25

Interviewer: *In operational situations, have you received pressure or influencing force to make a decision on a structural integrity risk?*

Interviewee: Yes.

Question 26

Interviewer: *In your opinion, this is a disagree/agree one, in your opinion in-service structural integrity risk analyses are used effectively by decision-makers.*

Interviewee: Say that one again.

Interviewer: *in your opinion in-service structural integrity risk analyses are used effectively by decision-makers.*

Interviewee: I think I agree. The current structure is used effectively to the extent that it can be by the people that are using it. Because people are trying their best basically with what we've got.

Question 27

Interviewer: *But bear in mind what we've mentioned about it requiring improvement. Ok. Considering situations that you have seen of structural integrity risk management decision making, do you believe decision-makers understand the assumptions and uncertainties presented to them from structural integrity risk analyses? So there's a scale here between never and always for both uncertainties and assumptions. So take uncertainties first, do decision-makers understand the assumptions presented to them.*

Interviewee: I think most of the time.

Interviewer: *And assumptions*[4] *?*

Interviewee: The same.

Question 27a

Interviewer: *Can you give any most common reasons that decision-makers might not understand assumptions with structural integrity risk analysis?*

[4]Noting the error in interview confusing the assumptions and uncertainties, the answers are identical therefore there is no requirement to discount the answers as void as a result of the confusing presentation of the question.

Interviewee: So particularly on *[aircraft type]* you're working with other nations so potentially you'd have language barrier issues where assumptions would be made on one end and potentially not transferred to the other. Going back a little bit to what we were saying before where you've got people like myself going to the PT for 2 and half years, where you've got people in the design organisation who are in that team for 20 years or whatever, they might assume that you've got a base level of knowledge that you potentially don't have, so they might make an assumption that you've kind of already read that across. And occasionally it's just oversight, which everybody is prone to from time to time, missing a line in a document that's actually got a significant....

Interviewer: *that's kind of related to the last point I suppose then isn't it where you might not get the relevance of it but somebody else would jump on it and go "oh crap what about this"....*

Interviewee: yeh you might skim over a line and not really acknowledge that it's really important, whereas somebody else might see that as being like "oh my god we really need to focus in on that and do some more digging".

Question 27b

Interviewer: *What about uncertainties?*

Interviewee: I think the same. I think pretty much as above.

Interviewer: *Same reasons?*

Interviewee: Yeh.

Question 28

Interviewer: *Any comments with regards to that section we just did there?*

Interviewee: Only that I think there's two types of risk assessment that we're talking about. There's kind of the operational "I've got a decision on one aircraft that I need to make a decision about" and "actually I'm make a decision for the fleet" and where we've been discussing some of the uncertainties or assumptions you've got more, it feels like you've got more time in those fleet decisions to probe that little bit further, so you're less likely to have those issues I think when you're looking at fleet wide bits because it tends to be that those longer term problems, you'll not only do over the phone, email and tech queries, but you'll probably end up being discussing those issues in person in either formal meetings and also in the sidelines of meetings as well, so I think that reduces the risk so, there's a lot to be said for those structured meetings.

Question 29

Interviewer: *So last section, and we're almost running to time....so this is all about the management of structural integrity risks, kind of the bigger picture. Were you previously, or have you currently been formally delegated the authority to own and decide upon structural integrity risks in your position?*

Interviewee: Yes.

Question 29a

Interviewer: *Did/do you own this risk solely or in conjunction with others?*

Interviewee: I'd say in conjunction with others.

Question 29b

Interviewer: *If you were/are not able to decide upon a particular in-service structural integrity risk, who was the ultimate decider?*

Interviewee: So pass it up to the *[senior aircraft airworthiness authority]*, or if it was a risk to life issue, then to the Duty Holder.

Question 30

Interviewer: *It's another disagree to agree bracketed jobby. Your organisation's risk management method enables you to maintain a transparent audit trail, which details the history of the treatment and ownership of in-service structural integrity risks.*

Interviewee: Agree.

Question 30a

Interviewer: *What measures/aspects of the risk management framework enable the process to be considered transparent?*

Interviewee: I would say that the *[hazrd management working group]* meetings that we had, the Structural Integrtgiy Working Group and the associated minutes that we had, the ...God what were the names of the meetings....where we sat the Duty Holder and informed them of the issues, we had aircrew and forward representation at the monthly *[hazrd management working group]*, so it wasn't as if we were just marking our own homework, we had the external people coming in. We also had the Indepdent structural advisor from *[other company name]*and we had *[other company name]* coming into the *[hazrd management working group]* as well to add that level of indepdencen to mean it wasn't just us marking our own homework.

Interviewer: *Ok so very meeting heavy, but the minutes of those meetings were critical to ensure that transparency.*

Interviewee: And the associated supporting loose minutes that were generated ahead of those *[hazrd management working group]* meetings that were available to all members and were certainly not hidden.

Question 31

Interviewer: *Who do you believe acts upon the risk analysis conducted for in-service structural integrity issues? You've got Operator/end user, regulator, manufacturer, equipment owner, line manager, another department internal to organisation....*

Interviewee: Can it be all?

Interviewer: *Yep. Any others on top of those, who might use the risk analysis?*

Interviewee: Future designers? Probably not.

Question 32

Interviewer: *To what extent does your structural integrity risk management strategy explicitly try to identify rare events, 'black swans' or...they're called black swans...have you heard that term before?*

Interviewee: No, but I know what you mean. Is it something that is unlikely to happen but.....

Interviewer: *Yeh...not.....It's something that would surprise you if it happened, unknown unknowns is the other... this is Not at all considered / To a very little extent / To some extent / To a great extent / To a very great extent...*

Interviewee: I think to some extent because we did do exploratory teardown...managed expertly I always thought by someone else in the PT.....but yeh so to some extent. I think in some respects that was because of the team that we were in. Had we been in a platform with fewer aircraft that weren't being retired, the level of that would probably have been reduced, although there might have been more done at servicing points.

Interviewer: *Do you think that was...so there was nothing necessarily in process that allowed for those unknown unknowns to be...*

Interviewee: There's the aging aircraft audits, so that would...and that's across all platforms. So that helps dig into unknown unknowns, so yeh there is that, but interms of the follow-on level, I think there is an element of that's down to the risk appetite of the *[senior aircraft airworthiness authority]* and the duty holder, also operational risk is taken into consideration so for example if you are going to take out one of three aircraft for 12 month in order to do exploratory stuff, that's obviously going to have a far greater impact than if you take out 1 of 120. So yeh I think it's all, all has to be again judgement-based, because if you start putting it down as "you must do a certain percentage" then actually for some fleets, you could actually end up crippling them where as others wouldn't even really notice it so, yeh.

Question 32a

Interviewer: *Are those formalised techniques, what you mentioned interms of the policy required, aging aircraft audits.*

Interviewee: Yeh.

Question 33

Interviewer: *Is Structural Health Monitoring planned, or currently in use on your equipment for use in Structural Integrity risk management?*

Interviewee: Yeh.

Question 34

Interviewer: *What methods are used to provide confidence in the risk analysis conducted (for example independent review, consensus of opinion, audit)?*

Interviewee: Certainly the first two. Audit I would say is not really. It's used but not audit the decision, it's more to audit the process used.

Interviewer: *Any other methods to provide confidence, in the risk analysis?*

Interviewee: No, we've got kind of lots of different waqys of doing review by peers and what have you, but it's essentially just discussion with others isn't it. There's lots of that, and lots of different ways of doing that, but no.

Question 35

Interviewer: *Is assurance activity carried out to validate the effectiveness of your structural integrity risk analysis? Yes, No, Unsure.*

Interviewee: Sometimes....i think it depends on the issue, I think if it's a really significant issues that's still on the radar, then, that's been mitigated but hasn't maybe been mitigated down to a level where it's considered to be low enough, then yes. But if it's something that you'd go "right that's done" put to bed, then unless the issue re-rose it's head 10 years later you wouldn't look at it again.

Interviewer: *So it sounds like it's still affected by judgement then?*

Interviewee: Yeh.

Question 35a

Interviewer: *Who conducts assurance activity? Is it self audit / internal audit (internal to organisation) / external audit (external to organisation, but within same industry) / external audit (external to organisation, but outside of your industry).*

Interviewee: Probably a bit of a mixture. So you do get independent people from outside coming in and doing audits, people like *[other company name]*, people like *[other company name]*, but to be honest most of the repeated reviews tend to be lead in house by the desk officers, going back over things and, yeh. And bringing things back up. I didn't see a lot of "right I'm going to audit something that has been considered closed and put to bed", I didn't really see that at all.

Question 35b

Interviewer: *Do you believe the assurance activity is effective?*

Interviewee: To some extent?

Interviewer: *It's a yes or a no!*

Interviewee: I think because I can't say "absolutely yes", then it has to be a no doesn't it? I think if there was a scale on that, then I would say "ish".

Interviewer: *More on the yes, or more on the no?*

Interviewee: More on the yes, because we do go over things at the *[hazrd management working group]*, things like we have those meetings where we go back over bits and pieces in the SIWG etc. so in terms of assurance, yes a lot of work is done to ensure that decisions aren't just made and put to bed. But it's only just on the yes, because for example if I made a decision on something and hadn't escalated it high enough to reach those meetings, it could easily not get looked at ever again.

Interviewer: *So it could easily be influenced by, again, your judgement?*

Interviewee: Yeh.

Question 36

Interviewer: *What changes might be made to improve your organisation's structural integrity risk management?*

Interviewee: what changes might be made to improve it.....made to the organisation or the policy, or both? I'm not sure I have an answer for either, just interested in the question!

Interviewer: *It's loose and open to your interpretation, so what would you change if you were to make risk management for structural integrity better within the organisation, what would it be? It doesn't matter if it's policy, there are no bounds to that.*

Interviewee: I really don't know how you would make it better.

Interviewer: *If you've got no answer, don't worry.*

Interviewee: I don't think I do because if I knew how to make it better, I would have tried it.

Question 37

Interviewer: *... Do you have any comments on the answers provided in this section?*

Interviewee: Just that the questions were excellent.

Interviewer: *...Thank you very much that was really useful.*

Interviewee code	171220192158 (ID 59)
Organisation industry	Aviation.
Organisation category	Operator/end user
Role when handling risk	Risk decision-maker

Interviewer: *Because it's anonymised it's 17 Dec at 2158, are you happy for me to record our interview?*

Interviewee: Yup

Interviewer: *thank you very much. So the ethical side which I'll just run through quickly.....[discussion on ethical aspects of study]....with that in mind are you happy to continue?*

Interviewee: Yup.

Question 1

Interviewer: *Thank you. The first section is....sets the flavour for. So within which industry is your organisation? I'm going to to go with aviation, I'm assuming?*

Interviewee: Yep, military aviation?

Question 2

Interviewer: *Which category best describes your organisation? Operator/end user / Regulator/policy maker / Designer/equipment manufacturer / other.*

Interviewee: The first.

Question 3

Interviewer: *What's your role? That's with specific regard to any time that you've been structural integrity, but I've had anything from engineer officer to structural integrity desk officer, to mechanical systems.*

Interviewee: So current role, reds and greens holder, level J on a front line aircraft.

Interviewer: *What about when you were in the [aircraft project team]?*

Interviewee: I had *[airworthiness delegation]* across the entire platform including structures.

Question 4

Interviewer: *In your experience, do organisational risk management policies specifically consider structural integrity risk?*

Interviewee: Yes. There are policies for it.

Interviewer: *Specifically for structures or is it held within a generic application?*

Interviewee: Do you know what, I think it's generic, but I think structural risk is the model it best fits. Things like ESVRE probably was conceived with structure in mind, but if it's thought process, yeh.

Question 5

Interviewer: *Within the context of equipment structural integrity, which of these best describes your role (or the role most usually perform) when handling risk? [Risk analyst / Risk decision-maker / Commentator / Advisor]*

Interviewee: Risk decision-maker makes the best fit for me, current role.

Interviewer: *Presumably that's applicable to the Project Team as well given you had [airworthiness delegation] across the platform?*

Interviewee: yeh.

Question 6

Interviewer: *Describe the current process for conducting risk analysis for in-service structural integrity issues, from identification/realisation of an issue to the point when the risk is considered managed.*

Interviewee: Oh dear, I'm probably too long out of specific courses to give the dictionary definition answers.

Interviewer: *I don't need dictionary definition, what do you do, what process do you follow?*

Interviewee: I suppose I should talk through how it would happen, someone would raise a defect, a structural defect, they would then inform, generally if I were involved in that management of aircraft that day, they would inform me of the defect potentially that they were working it, or if there was going to be a length of time required to work it they might ask whether or not there was sufficient time to do that before pulling the work. Lets say they couldn't do the repair, or we elected not to do the repair we would then have full measurements taken, some elements of structure require NDI, some require paint or surface finish specialists to strip back in order to get measurements and so on, photogs maybe as well. We then look to Technical Orders to establish whether there are limits established for that piece of structure, either specific or generic, because of being primary, secondary, tertiary etc. If there were limits, measurements would have to be inside limits, and then it would be fairly straightforward for the fault, ADF [acceptable deferred fault], if they were outside of limits may still defer the fault but do so in accordance MAP 5.4.2 raising the necessary paperwork with the associated time limits, which I think is 7 days for outside limits from memory is it ... if we felt that contract support from the OEM or DO was required, or advice on the damage and it's acceptability or a repair scheme was required, we would raise our version of a technical query and that would go off via the system that we use. Technical queries come back to us, or we have sight of them as soon as the OEM has finished with them, and we can use them under our own authority but we would generally wait unless it was something fairly benign or we needed the aircraft, we would wait until the DT *[delivery team]* or *[airworthiness delegation]* holder issued their approval as well to make it approved data.

Interviewer: *If it went outside the 7 days while you were waiting for that information back, what would you do then?*

Interviewee: Probably re-defer it depending on the damage, if it was something I was comfortable re-deferring if it was tertiary, or non-pressurised, or non-loading bearing, or non-structural, then fairly easily. If it were something that I felt were primary, I probably wouldn't have ADF'd it in the first instance without the advice, but to be fair the advice that we get from our OEM is 36 hours is their contracted reply time, so we're unlikely to be in that scenario, and *[inaudible]* properly thought it through.

Question 7

Interviewer: *I appreciate it's difficult to do it in a general sense without having a specific example in front of you, so I appreciate the different options you've given, and don't worry about if you do name a company, that's all anonymised as well, so it won't be recorded along those lines, ok thank you. There's several questions throughout that will be on a Likert scale from strongly disagree to strongly agree, so if I say it's a Likert answer between those two*

then if you just intimate where on there and I'll give you the options I've got rather than go through the options I've got.

Interviewee: So strongly agree, agree, neither, disagree and strongly disagree.

Interviewer: *There's two more in there, there's strongly disagree, disagree, slightly disagree and then the neutral point, and then the mirrored for agree. So this phrase is whether you strongly disagree or somewhere on that scale.... In my organisation, structural integrity risk analysis is an activity that is valued*

Interviewee: Strongly agree.

Question 7a

Interviewer: *Why do you believe that to be the case?*

Interviewee: It's current and our bread and butter, so it's not something we do as an exact science I accept that until it goes into FSR-type territory where they can do the exact science, something the air force generally can't any more, but it's something people take very seriously, you're not going to get people generally taking untoward risks when it comes to structural integrity or structural damage.

Question 8

Interviewer: *Within your organisation, are structural integrity risks that arise in-service described numerically or non-numerically/verbally/descriptively?*

Interviewee: As in measurements?

Interviewer: *The risks themselves, if the risk arises in service, do you slap a number on it or do you describe...*

Interviewee: On the first line there isn't a risk number associated to it. If I was in a DT *[delivery team]*, you may enter it into some form of risk management matrix and give it a score, yeh, we did use that on the DT we had that certain risks were given a risk matrix output.

Interviewer: *So I've put both.*

Interviewee: Yeh.

Question 8a

Interviewer: *How would you rate the accuracy of your organisation's quantitative risk analysis of structural integrity? And this is between very poor and very good.*

Interviewee: If it's just my organisation, then probably middle of the ground, it's not brilliant but it's not really bad, it's fairly qualitative rather than quantitative, but you asked about quantitative so probably very neutral

Interviewer: *The option I've got is very poor, poor, fair, good, very good, so middle of the road is fair?*

Interviewee: Yeh, but when you're talking of the entire organisation in that we have support of FSRs, OEM, DO and all these kinds of people, we can and we do go back to the proper brain boxes in other places that can do full stress analysis using, what do they call them, computational

Interviewer: *FEM?*

Interviewee: Yeh that's it, that kind of thing, then it becomes very good.

Interviewer: *So it gives you a number that you feel you can trust more?*

interviewee Yeh.

Interviewer: *But you're not necessarily sure how accurate that might be?*

Interviewee: Er...well, take an example recently, we had an actuator support bracket that had cracked, and they did the full analysis of the full load through that under different conditions of loaded, or in flight, or different attitudes and configurations of flight, and worked out the stresses, worked out all the..everything about it, so it gave us a really concrete answer as to here's what it can take, here's what a repair needs to look like, here's how big it needs to be, everything, yeh, it was very mathematical, although the maths doesn't matter too much because we don't actually use that we just cut the metal and stick it on, but we have the background behind all that.

Question 8b

Interviewer: *So you mentioned there were sometimes when you verbally describe the risks, what's the reason you don't quantify the risk in those circumstances?*

Interviewee: Fairing, panels, things like that, it might be a verbal there's no risk of this panel coming off, if it does come off, it's rear of the intakes, or it's not going to cause a FOD issue.

Interviewer: *So it's much more consequence based rather than saying the likelihood of it happening?*

Interviewee: Yeh, but that's generally where you're content that it's not load bearing. Where it's load bearing you'd want to do that maths, you'd want that FSR or like *[specialist structural repair organisation]* what do they have the accreditation that means they can sign off

Interviewer: *DAOS?*

Interviewee: Yes, you'd want that sort of level of approval someone who can do the maths. If it's not load bearing, that's where you go into that verbal description of...36 of 38 fasteners still remain...

Question 8c

Interviewer: *In my opinion, a quantitative, numerical, value is necessary for the assessment and management of in-service structural integrity risk?*

Interviewee: When we say quantitative, are we talking about the analysis of the actual load figures of that piece of structure, or are we talking about a score assigned to the risk of that failing?

Interviewer: *Either, so it depends on how, if you can convert from a number given for something, like you said about 36 of 38 fasteners, you can convert from that into a likelihood a lot easier than you can from starting with a blank canvas because you've then got some analytical data but it's down that route basically, do you need that, well it's how far you agree with that statement that a quantitative value is necessary for the assessment and management.*

Interviewee: If I can separate it to begin with then you tell me if that works. So for the actual structural integrity of a specific item, the numerical element, understanding loads stresses etc. is critical; for associating a risk of failure to give a score that is then aggregated to give a cumulative structural risk score, it would definitely add value, I don't think it is essential and it would also need a bit of a change in the way the MOD/Defence Aviation currently works, because we don't score everything in the same way. I think there are somethings we do, a lot we don't, you'd have to do it all to make it worthwhile.

Interviewer: *So you think basically risk management system is mixed in it's units or in it's composition?*

Interviewee: No I think that stress analysis is absolutely important, absolutely vital to operating the aircraft, scoring risks we don't do that really, not for structural stuff and we'd need to change quite a lot to get there. So the aircraft are not safe to fly at all without doing that structural analysis. The aircraft can still safely operate without numerically working out all the different risk scores and every piece of structural, every structural fault. However it would definitely make it safer, it doesn't make it unsafe to fly without it, because we don't have it ay the moment, so it would definitely be a step safer.

Interviewer: *That's how we worked at [aircraft type team], but in the world of risk that we operate under RA1006/1012 we're expected to demonstrate safety to a combined level of* $1x10^{-6}$ *loss per flying hour.*

Interviewee: Yeh true, That's more a type concern at the moment rather than continuing. I think we don't really do that in the continuing space is what I'm getting at. You get a figure from the DO saying here's the likely of your aircraft crashing for structural failure, it's $1x10^{-6}$ or less, brilliant we're safe, we've got that from the OEM and accredited by the ITF or whatever they're called. For continuing airworthiness we've got this crack, we don't feed that in a model of what's our effect on structural health now numerically, not that I'm aware of. That would add value.

Question 8d

Interviewer: *[Discussion about motivation for the research.] Ok so related to that in a similar statement for strongly disagree/strongly agree somewhere on that spectrum, "In my opinion, a qualitative (verbal/descriptive) interpretation is necessary for the assessment and management of in-service structural integrity risk?"*

Interviewee: Strongly agree I think, I'll go with that.

Interviewer: *So I've put strongly agree for both of those but I've also put "caveated" so I'll annotate what we've talked about.*

Interviewee: So Strongly agree because I think a description of the risk is really important.

Question 9 and 9a

Interviewer: *When handling in-service structural integrity risks, are you required to assess structural integrity risk relative to a threshold level?*

Interviewee: I can't think there being a defined threshold, other than at the subsequent enquiry "ALARP", I can't think if there's any other limit of how much you need...can you ask the question again sorry? *[Asked again]* At my level I don't believe I am, it'd be interested to know if I am, is there an RA that I says I should do that, there is.

Interviewer: *so it depends where you believe ALARP sits, I think, that's my opinion.*

Interviewee: I think that's where I'm getting at, I think the only definition... the only standard I need to meet is ALARP and were it ever to go wrong I could defend myself.

Interviewer: *and interestingly the ALARP statement drags in the likelihood and consequence risk, which is a quantified value in our current state. So what I've put there is I've put Yes because you've mentioned ALARP, which is an interesting answer. Well it's not one I've seen before, but I think it's absolutely valid. Are you happy with me putting that by the way?*

Interviewee: Mmmm yes it's perfect.

Question 9b

Interviewer: *In operational situations which are information poor, time critical, or resource constrained, I believe a threshold level is appropriate. And this is a strongly disagree to agree. An operational situation labelled here is one that's time critical, information poor or resource constrained.*

Interviewee: Can you ask the...start the question again sorry?

Interviewer: *Yep this is... In operational situations which are information poor, time critical, or resource constrained, I believe a threshold level is appropriate.*

Interviewee: "a" or "the" threshold level

Interviewer: *A threshold level.*

Interviewee: And would that be a numerical threshold?

Interviewer: *Whatever the threshold level that you think is....so to put it into normal speak, if you're in an operational situation and you have to have a pass/fail level, basically.*

Interviewee: No, so I would disagree, not strongly, but I would disagree, that is the whole raisen d'etre of the military airworthiness reds or greens holder, however caveated that with operating risks vs operational, operating risk is my domain as an engineer, operational is the air component commander or the *[duty holder]*, whoever, so if I don't sign that ADF or if I don't agree to that aircraft to fly unless the aircraft is safe, however for the troops in contact, or the critical spare, or the munition, or the whatever, the desired effect, which the operational element is lost, I think you need to have that flexibility.

Question 9c

Interviewer: *Who sets, or how is that threshold level set?*

Interviewee: At the moment it's very personal isn't it, if something goes wrong.

Interviewer: *Quite and then who does it fall back on.*

Interviewee: The regulator I suppose?

Question 9d

Interviewer: *So I'm going to list a load of processes and if you just say yes to the ones that you use in support of structural integrity risk analysis or decision making. So... if you've not heard of them then I can explain but if you haven't heard of them then the likelihood is that you don't use them. What if analysis?*

Interviewee: Yes.

Interviewer: *Check lists?*

Interviewee: Yes.

Interviewer: *Safety audits?*

Interviewee: Yes.

Interviewer: *Task Audits*

Interviewee: No.

Interviewer: *Bow tie?*

Interviewee: Yes... is... bow tie is only just creeping in for engineering, bow tie is very much an operator thing with a sort of ****ing about in the edges trying to get engineering, but again heard of the principle but haven't seen it applied to structure. I can't think of an example where it's been applied to structure.

Interviewer: *PRAT?... it's a specific technique, I haven't seen it used in the air force.*

Interviewee: In which case I'll say no then.
Interviewer: *Risk matrix?*
Interviewee: Yep
Interviewer: *Quantitative Risk Assessment, you'd described that you'd seen it used?*
Interviewee: I have, more in DT *[delivery team]*.
Interviewer: *Statistical classic approaches: probability distributions?*
Interviewee: Yep, RCM type stuff.
Interviewer: *Event data models, using historical data and frequency of failure?*
Interviewee: Yeh, so again RCM 5A1 for an aircraft is very much just that.
Interviewer: *Human error analysis technique?*
Interviewee: As a specific technique not sure if I've heard that term, but as a concept yes.
Interviewer: *It's a specific technique, it's looking at human processes and human failures in the system.*
Interviewee: Generically most certainly but I'm not sure I've heard that exact name so what would that be would you say yes or no, I haven't heard the name so take it off.
Interviewer: *So things like [error management system] might be considered within that type of area but [error management system] is more post-event where as human error analysis technique is more, you're looking more proactively at how someone might make a failure, or a failure of judgement.*
Interviewee: I mean when I've been involved in scoping new equipment ground support equipment, HFI human factors integration was one of the steps, one of the processes to go through for the approval of that, for the validation/verification, so HFI was an element of that and you're looking at is the person going to hit the big red button rather than an attendant button, the shutdown sequence
Interviewer: *It's a group of techniques*
Interviewee: I completely understand what the words seem to mean to me but that specific title I would have to say no.
Interviewer: *Specific to structure, if you've not seen it in structure then don't worry.*
Interviewee: Personally I haven't no.
Interviewer: *Fault Tree Analysis?*
Interviewee: Yep
Interviewer: *Event Tree Analysis?*
Interviewee: Yep.
Interviewer: *Risk Based Maintenance?*
Interviewee: erm, no, erm sort of, I don't think I've heard of the technique but it is what RCM is when they say over the past 20 years this has failed 5 times, therefore it's going to fail once every 4 years, so we'll change it every 3 years, so yeh that's.
Interviewer: *Epistemic models?*
Interviewee: Mmm no.
Interviewer: *Bayesian Networks?*
Interviewee: What's that

Interviewer: *It's a fancy technique, we definitely don't, it's taken me a year to get my head around it, not saying that's not a reason to introduce it, I think they're very good, but yeh. Stochastic processes, monte carlo simulation*

Interviewee: Yeh, I've seen Monte Carlo, where've I seen that? I've definitely seen that, but I'm not sure I've seen it in structures.

Interviewer: *Bayesian Analysis?*

Interviewee: You've mentioned that one.

Interviewer: *They're different.*

Interviewee: I thought you said Bayesian.

Interviewer: *Bayesian Networks and then this is Bayesian Analysis.*

Interviewee: So Monte Carlo yes but the other ones no. In fact no, I've not seen Monte Carlo in relation to structure, what's the one I have seen...FMECA?

Interviewer: *Don't know, it's sounds like a programming language to me*

Interviewee: No it isn't its like fault something....

Interviewer: *Ok don't worry about it, we've still got a few questions to go so I'd like to keep moving...*

Interviewee: Failure Mode and Effects Criticality Analysis

Interviewer: *Oh F.M.E.C.A, no I've heard of that, yeh, so that's like a fault tree ish, it's, if you've seen that done then I'll put it in the "other (specify)".*

Interviewee: Yeh I've seen that done.

Interviewer: *Probabilistic distributions for key variables*

Interviewee: Again is that your kind of failure curves, bath tub, infant mortality, that kind of thing?

Interviewer: *It's kind of like taking in an analytical sense you say that your crack propagation rate has this type of distribution, it's most likely to be at this value but it could also be this value, or between this value and this value*

Interviewee: I probably don't get to that level of depth.

Interviewer: *Ok. Bespoke numerical models for a given situation?*

Interviewee: Yeh.

Interviewer: *Neural networks?*

Interviewee: I think so, but remind me again

Interviewer: *Things like machine learning, using training data sets to predict how things are going to perform.*

Interviewee: Yeh, obviously I didn't do but that's what the RCM guys that I worked with were doing.

Interviewer: *Any others that I haven't specified?*

Interviewee: No, only that FMECA one, I think.

Question 9d1

Interviewer: *Do you believe these processes that we've just gone through are appropriate for structural integrity risk analysis?*

Interviewee: The ones that I know of, to the best of my limited knowledge, yep, but they each have their place.

Question 9d1a

Interviewer: *And why the reason for that is? You've just said they've each got their place, unless you want to add anything else?*

Interviewee: As a non-structural, what I mean by that is someone that can work out things like crack propagation rates through critical fracture joints, all that kind of stuff, I left all that behind 10 years ago when I left university, but from my understanding those kind of overarching models work.

Question 10

Interviewer: *Is your organisation's risk analysis methodology proactive, enabling anticipation of risk from in-use changes or modifications to usage/configuration, or is it reactive, which is responding to issues that occur?*

Interviewee: Mixture of both. Condition based vs predictive maintenance isn't it, most of our maintenance is condition-based.

Interviewer: *I mean both of those, preventative is pre-planned, but what I'm trying to get at here is is there anything where you sit down and you go well what could happen?*

Question 10a

Interviewer: *What methods do you use to identify risks?*

Interviewee: Things like monte carlo, we've done that one. For structural stuff I'm trying to think.

Interviewer: *So how did you use monte carlo simuluation?*

Interviewee: I think it was more for, I think I did it on my acquisition course, trying to think what we used it for... might have taken it off the list..

Interviewer: *Yeh you took it off the list.*

Interviewee: ...it's already gone has it? Right so, I've done risk management course, I've done acquisition course, and I've done the, what's it called *[aircraft structural integrity course]*, I'm thinking more about the *[aircraft structural integrity course]* and trying to remember what risk analysis, not analysis but risk generation processes we used. I think the only time we'd have done that is operating-type risks, so in-flight, I've been involved in that. Structural health type thing I'd expect that to be more OEM/DO, maybe ITEAC.

Interviewer: *So it's not necessarily something that you guys have done?*

Interviewee: No, not structural health, not structural integrity. Clearly it is done, and the RCM guys would have done it based on historical data as well and how many, I don't know, axle failures happened or whatever brackets fails over a certain period and feed that into their algorithms to validate your maintenance package, or introduce maintenance if there isn't one before and it's been conditional. So that's fairly reactive actually, we have done that, we have done it where it's been based on historical data or based on events, and we've introduced structural health management processes where previously it had been conditional, and we've said right here's an *[technical instruction]*, or a change to the *maintenance schedule* to inspect for cracks, it's more reactive but created something more proactive.

Question 11

Interviewer: *The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'likelihood' of occurrence. Do you strongly disagree to strongly agree?*

Interviewee: I would go for the neutral one, because it's the best we've got but it's not great, it's all just....i don't think you can get it much better but it isn't great, you're trying to predict the future and no-one can do that.

Question 12

Interviewer: *The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'consequence' of occurrence.*

Interviewee: I think is it, I'd go one up from neutral on that one.

Interviewer: *Slightly agree?*

Interviewee: Yeh.

Question 13

Interviewer: *More often than not, there are uncertainties within my analyses of structural integrity risks. So uncertainties here are information or a value that is not known accurately.*

Interviewee: Agree yep. Between strongly and agree, we'll go for agree but that's why we use risk factors and correction factors.

Question 13a

Interviewer: *Are uncertainties communicated in risk analyses regarding operational structural integrity?*

Interviewee: Generally yes, so the ones' I've done before now are when we've looked at mod, done what they call them CBAs *[cost benefit analysis]* of the mod program for a mod for a particular structural failure, we fed in likelihood of occurrence and we used that to work out whether to fund the mod or not, but I think that's doen to a much greater level in the OEM/DO.

Question 14

Interviewer: *More often than not, there are assumptions within my analyses of structural integrity risks.*

Interviewee: Strongly agree.

Question 14a

Interviewer: *Are those assumptions communicated?*

Interviewee: urm agree (that's a yes or a no), oh yes!

Question 15

Interviewer: *More often than not, I handle structural integrity risks conservatively. So conservatism here is considered as cautionary thinking, over-estimation of the risk resulting in additional measures or values added to provide a safety factor given uncertain quantitative outcomes.*

Interviewee: Yeh, I think slightly agree even agree. Go for slightly.

Question 15a

Interviewer: *When conservatism is applied, is it communicated in risk analyses regarding operational structural integrity?*

Interviewee: Yes.

Question 16

Interviewer: *To what extent do you believe human subjective judgement (from expert opinion or belief) influences the analysis of in-service structural integrity risks?*

Interviewee: To a great extent.

Question 16a

Interviewer: *Are human subjective judgements communicated in risk analyses regarding operational structural integrity?*

Interviewee: Is that a yes/no? Yes, not the strongest yes, but I'll go with yes.

Question 17

Interviewer: *I am confident in the structural integrity risk analysis method applied for handling operational risks.*

Interviewee: Slightly agree. I have some confidence in it but not a huge amount.

Question 17a

Interviewer: *Why's that?*

Interviewee: Because I mentioned that analysis process of yes I'm happy that there's 36 fasteners instead of 37, I feel that my application of that rule or that ability to defer faults is different to others. And I have seen an ADF fairly recently where someone said "oh yeh there's still 9 fasteners left out of the 10 in that" but these fasteners were like 3/4 bolts holding supporting several tonnes of a moving component, and I thought why on earth they had made that decision.

Interviewer: *So there was a lot of individuality about it basically?*

Interviewee: yeh.

Question 18

Interviewer: *Do you have any other comments on that section?*

Interviewee: Nothing at all.

Question 19

Interviewer: *Section three [introduction]. What information is essential in order to make a decision upon an in-service structural integrity risk?*

Interviewee: DO detail.

Interviewer: *Like their sign-off, their assessment?*

Interviewee: No sorry TO, Technical Orders. So obviously what component is it, identify it in the IPC/IPB, and then with that whether there is any limits for damage, is there a repair scheme in the *[structural repair manual]*, or that kind of thing, and then from that, any measurements, NDI, approved techniques for NDI. What else do you need to know. They're the initial need to knows for some faults, you could probably defer based on that, along with the time it would take to repair, whether the repair is within our capabilities, they're the factors that I, the things that I factor in based on whether I'm going to make a decision to defer a fault or not.

Interviewer: *Wider maintenance considerations like resources and things?* Yeh any resources like time, people, specialists etc. And then if it were to be deferred and I was to go to the OEM/DO for advice, I would want their structural assessment for that I imagine measurements or whatever.

Question 20

Interviewer: *So the next question is describe the decision-making process regarding in-service structural integrity risks.*

Interviewee: Yep, so identify it get the tech docs, take full measurements and description, is it in limits, yes, cool, we'll probably just crack on. No, possibly crack on, and then asking for advice, or I'd ask for advice *[inaudible]*, or if you can just fix it straight away then just fix it, or repair it, or replace.

Question 21

Interviewer: *Do you ever use judgement when making or supporting a decision regarding in-service structural integrity risks?*

Interviewee: Yes.

Question 21a

Interviewer: *In which types of situation do you use judgement? And there's a list here so if you just stop me when it applies. When there is a lack of available expertise?*

Interviewee: Yes.

Interviewer: *When time constrained*

Interviewee: Yes.

Interviewer: *When I disagree with advice*

Interviewee: I can't think of a specific example just now but yes I would probably would.

Interviewer: *When there is insufficient resource to support a decision choice that would otherwise be available*

Interviewee: Yes

Interviewer: *When there is insufficient information available to support an objective analysis*

Interviewee: It depends for that one

Interviewer: *Upon what?*

Interviewee: It wouldn't be my first preference to make a judgement if I didn't know a lot about it, it's a case of "ooo I think that's alright" then as that's what the question sounds like then I'd be less inclined.

Interviewer: *When I believe that a particular decision choice is 'good enough'*

Interviewee: Yes

Interviewer: *When I believe that technical information is conservative.*

Interviewee: For structural integrity?

Interviewer: *Yes*

Interviewee: I might but probably not, but seeing as it's a binary yes/no, let's go with no.

Interviewer: *Any other situations when you might use judgement*

Interviewee: I think that's covered them all.

Question 21b

Interviewer: *When you use judgement is it holistic, using your gut intuition or instinct, or do you decompose your judgement upon the situation into constituent parts.*

Interviewee: I would like to think that I'm more analytical. There is definitely an element of personal opinion, but hopefully that is backed up by training and experience that has been to a reasonable standard and has been proven. Yeh ok, so what are the options again.

Interviewer: *Holistic or decomposed?*

Interviewee: Do you know what, I think my gut is that I think about it a lot more. Do you know what I mean. Like my gut is not just, like I definitely wouldn't be going yeh I think that's fine go for it, it would be lets think about this properly or that is definitely just aerodynamic fairings because it says so in the TO. So yeh, even that's, what's it called?

Interviewer: *Decomposed.*

Interviewee: Yeh

Interviewer: *So decomposed?*

Interviewee: Yeh, there's a whole lot of instinct and things involved in knowing what is and isn't reasonable damage or problems or faults or whatever.

Question 21b1

Interviewer: *Are the judgements you make numerical or non-numerical/verbal? Or both.*

Interviewee: Well numerical if there's limits and things set, but non-numerical if it's the classification of structure, that's probably the biggest thing isn't it. You know, if it's internal structure, if it's just the floor in the flight deck, I'm probably not going to give it a huge amount of thought. You know, I will think about it of course, but I'll not be worried about structural integrity, if it's pressurised skin or wing spar or something like that, or flap or hinge, then I definitely would be going into the books, and then taking measurements, looking for classifications, so kind of both. The holistic thing is the first step isn't it and that's partly how you know whether or not to go into the decomposed.

Question 21b2

Interviewer: *When you decompose it, how do you represent the decomposed the situation?*

Interviewee: As in how would I then describe it, well it would be in the ADF wouldn't it. I'd follow the guidance for deferring the fault.

Interviewer: *So that's more getting at, so if you've got a crack in a frame, how do you go about, if you're saying you decompose it, then how do you go about extrapolating that circumstance on your judgement given you haven't got any analysis infront of you.*

Interviewee: Looking at whether or not this was primary or secondary, a frame I would not be signing off myself, not until I had something from a structural specialist to tell me "here is the implications for that frame" unless there's set limits.

Question 21c

Interviewer: *What do you find difficult when using judgement to support decisions regarding in-service structural integrity risk?*

Interviewee: Generally when I'm away down route or it's an operational pressure, that makes it more difficult. What do I find difficult...I don't think I do find it especially difficulty, there's process and there's thought, it's no more difficult than any other tasks that I do, it's just it is what it is, you know what I mean?

Question 21d

Interviewer: *Yeh, I'm trying to look for any clues that anyone might give in terms of if there's something that's really vexing about trying to make a call on a structural integrity issue, and then related to that is the next question which is how could you be supported better when using judgement to support decisions regarding structural integrity risk? If that helps?*

Interviewee: Well better classification of structure to say "this is important structure or unimportant structure with regards to the structural integrity of the entire airframe", generally that's inherent, but lets say for a floor panel this is not structurally significant, not for all of it, or does it? I think for the ones that it does, so for some of the cargo floor panels that I work with, I think they are part of the structure of the aircraft and they are marked in there. I think when you know something is part of the overall structural integrity of the aircraft, that's when you're much more inclined to go to the OEM and say...

Interviewer: *Get further advice?*

Interviewee: Yeh go for advice.

Question 21e

Interviewer: *Have you received training in applying judgement to support decisions regarding structural integrity risk?*

Interviewee: Yeh.

Question 22

Interviewer: *In operational situations, decisions regarding a structural integrity risk are clearly justified. Do you Strongly agree to strongly disagree that statement? [Question repeated]*

Interviewee: Do you mean that it's justifiable to make decisions for operational reasons or do you mean that the system is very clear in why those decisions are made?

Interviewer: *When you make a decision, are you able to clearly justify it in an operational situation?*

Interviewee: Slightly Agree.

Question 22a and 22a1

Interviewer: *Is there a mechanism for ensuring that structural integrity risk analyses are justified?*

Interviewee: Not sure really, other than deferred fault and something could go wrong, just defending it.

Interviewer: *But there's no specific mechanism in place?*

Interviewee: For what's it referring to?

Interviewer: *For ensuring that those risks, structures risks are justified.*

Interviewee: Lots of processes that we can use, but I don't think there's anything that says "you are to" "you must" or the aircraft can't fly without you doing this, just guidance really or a regulatory input.

Interviewer: *That kind of sounds like a yes, and then if I just list everything you've just said. Because there would be considered mechanisms, policies and things like that that you've just listed?*

Interviewee: yeh ok, so there's policy and regulatory articles, so I guess yes, correct. You could still circumvent, is what I'm trying to get at. You could say "aircraft is flying anyway", but there is policy there.

Question 23

Interviewer: *Are you more comfortable with numerical or non-numerical/verbal/descriptive representations of risky decisions regarding in-service structural integrity?*

Interviewee: I think I can work with either but I'm probably more comfortable with descriptive.

Question 24

Interviewer: *Within your organisation, are decisions upon structural integrity risk assessments "risk-informed" or "risk-based"? So Risk-informed is "Risk assessment results support decision making but are not the only consideration." And Risk-based is "Risk assessment results form the sole basis for decision making."*

Interviewee: The former

Interviewer: *Risk informed.*

Question 25

Interviewer: *In operational situations, have you received pressure or influencing force to make a decision on a structural integrity risk?*

Interviewee: Yes.

Question 26

Interviewer: *In your opinion, in-service structural integrity risk analyses are used effectively by decision-makers? Strongly disagree to strongly agree.*

Interviewee: Slightly agree. I've seen examples of very good and very poor, probably more good so just go slightly.

Interviewer: *Ok, so just describe why that might be the case, just expand that a bit more?*

Interviewee: I think I go back to the example there of the individual who decided to sign off on some quite unexplained and quite serious damage on a critical piece of equipment that failure could have had potentially quite a significant effect to her rest of the airframe, thankfully not flight safety but would have jeopardised the aircraft.

Interviewer: *and you disagreed with that?*

Interviewee: Yeh.

Interviewer: *so your judgement was fundamentally different from his?*

Interviewee: Yeh. I've seen some good examples as well...oh no I'm struggling, one would be I went down to recover an aircraft that was overseas that had some fairly severe structural damage, the *[internal UK MOD organisation]* guys came down and carried-out a temp repair, but with all the analysis behind it underwritten by their DAOS accredited repair...what do they call them...yeh, you know the guys that can sign off on a repair and say "that's a fit repair scheme" from *[internal UK MOD organisation]*? DAOS approved is that what they're called? This stuff looked funky because there was still quite a bit of damage to this structure, but all the analysis said yeh that's fine, so I signed it, it was in my name, you know, when we flew it back it seemed the process worked.

Question 27

Interviewer: *That's good, it's very good, that's a good example. Considering situations that you have seen of structural integrity risk management decision making, do you believe decision-makers understand the assumptions and uncertainties presented to them from structural integrity risk analyses? So this is broken into two where you've got an answer for both assumptions and uncertainties.*

Interviewee: Can you read the question more slowly, sorry.

Interviewer: *[Question repeated] This is a scale between never and always.*

Interviewee: I'd probably say most of the time, or I think some of the time really.

Interviewer: *So "sometimes", "about half the time", "most of the time". So "sometimes" is between never and about half the time.*

Interviewee: About half the time? That's being a little bit pessimistic.

Interviewer: *Is that both of them? For both uncertainties and assumptions?*

Interviewee: Uncertainties about half the time, assumptions one up from that, what's that?

Interviewer: *Most of the time.*

Interviewee: Most of the time yeh.

Question 27a

Interviewer: *Are there, or are you able to list any common reasons that decision-makers do not understand assumptions with structural integrity risk analysis?*

Interviewee: Assumptions, I think they generally do, it's most of the time they understand assumptions. But perhaps the implications of the assumptions may not be fully understand.

Interviewer: *Why might that be?*

Interviewee: So you know someone assumes that that thing has been measured correctly and if it hasn't then that's a bigger impact, someone assumes that that thing has been manufactured to the correct standard, or assumes it'll be ok. So perhaps *[inaudible]*

Interviewer: *So is it a comms failing then?*

Interviewee: I think it's sometimes down to, people not being that good at their jobs?

Interviewer: *Are they perhaps not confident to be able to explain their assumptions?*

Interviewee: some people are overly, have too high risk appetites sometimes. They aren't diligent. They either don't have the wherewithal to be really diligent about it, or know when to ask for advice. Some people though actually are probably far too keen to ask for advice and too unwilling to exercise their own responsibilities.

Question 27b

Interviewer: *What about uncertainties? Are there any common reasons that decision-makers don't understand uncertainties?*

Interviewee: I think because exactly that, it's uncertainty so you're never going to know, so that's when you...when you have uncertainty, it's probably when you're going to go back for advice so you're not going to gain any more certainty from yourself you'll have to go to someone else.

Interviewer: *Are there any things like perhaps a lack of specific knowledge*

Interviewee: where not structural specialists, we're engineering managers, professional engineers as well, so have a good understanding of things but where things are uncertain, if there's a TO and it says a limit, there's your certainty, you know it's within limits, brilliant, assuming the person's measured it correctly and everything else about that is accurate, but, and that's very straightforward. But if it's outside limits, and you know that it is a piece of structure that has an effect on the airworthiness of the aircraft, you're uncertain because you're not an expert yourself, you're not the one who can do the maths. You might be able to and a lot of our guys I'm sure could, and you or I, if we spent the time learning, could but we don't, we've not been trained in that. So, probably we're not the right people to create those certainties. Does that answer the question?

Question 29

Interviewer: *Yep, pretty much, that'll do. I'm conscious I'm running time now so I'm trying to press you to try and keep us on, because I know you have a long shift and I don't want to keen you up. [Discussion] This section is about risks that arise in service and the handling of the analysis conducted on them. Were you previously, or have you currently been formally delegated the authority to own and decide upon structural integrity risks in your position?*

Interviewee: Urm, yeh.

Question 29a

Interviewer: *Did you or do you own this risk solely or in conjunction with others?*

Interviewee: Well some risks I own, others I do not.

Question 29b

Interviewer: *Ok both, situation dependent I'll say. If you were/are not able to decide upon a particular in-service structural integrity risk, who does make the ultimate decision?*

Interviewee: This is talking more about being in a DT *[delivery team]* is it?, is that kind of what this is getting at?

Interviewer: *Either, if you're where you are now and you're not able to decide on it, what happens. Or perhaps in the DT where did it go? Is there a next level up or...*

Interviewee: From me it's the DT. And when I was in the DT myself as an *[airworthiness delegation holder]*, I might pass it to the *[senior team manager]*, I might pass it to the *[other senior team manager]*, probably not really because his understanding and ability was, and delegation was the same of mine. Others may well go to the *[senior aircraft airworthiness authority]*. So yeh, there's escalation in terms of hierarchy.

Question 30

Interviewer: *So this is another strongly disagree to strongly agree. Your organisation's risk management method enables you to maintain a transparent audit trail, which details the history of the treatment and ownership of in-service structural integrity risks.*

Interviewee: Ohh no. Is that a yes/no or is it a scale?

Interviewer: *That's a scale between strongly disagree and strongly agree.*

Interviewee: I mostly disagree.

Interviewer: *You've got disagree, slightly disagree or strongly disagree?*

Interviewee: Disagree. Sometimes it is there, other times it isn't, probably more of the time isn't.

Question 31

Interviewer: *That's a really good answer thank you. Who do you believe acts upon the risk analysis conducted for in-service structural integrity issues? Is it the operator/end user, the regulator, the manufacturer, the equipment owner, the line manager, another dept. internal to the organisation, or any others that you might want to specify.*

Interviewee: It's probably all of those to some extent, so things like *[project safety meeting]* or whatever they call them on different platforms, you have all of those represented don't you so, to some extent all of them, but ultimately the *[senior aircraft airworthiness authority]* and *[duty holder]* are the biggest stakeholders in that I guess. The regulator doesn't own any risk do they?

Interviewer: *Ok do you want me to take the regulator off?*

Interviewee: Well yeh because they don't own any risk at all.

Question 32

Interviewer: *To what extent does your structural integrity risk management strategy explicitly try to identify rare events, 'black swans' or 'unknown unknowns'? A Black swan is a surprising event that may not have been considered when the decision was being made, but that would have changed the decision if it had been considered.*

Interviewee: So on my level in a forward environment, probably not a question I can answer very much, although it probably is in individual examples of specific structural problems.

Interviewer: *What about in the [aircraft project team] then?*

Interviewee: In the *[aircraft project team]* we outsourced that really so it's difficult to say. We outsourced that like to people like RCM, the contractors, or the OEM. Like for operational risk we did a lot of the bow tie and urm...

Interviewer: *That's addressing risks that... I know it sounds stupid... that you know about. Do you recall any specific examples where you actually sat down and went, right, lets go through and lets consider things that we haven't considered.*

Interviewee: Yeh so we did do that, again it was more about operating the aircraft so when the aircraft is in the air, what might a bird strike lead to, what might an overtemped engine lead to, what might an engine failure after take-off or a main landing gear fail lead to, those sorts of questions, we'd go through those and analyse them, but it wasn't with a structural integrity focus, it was with an operational risk focus, or operating risk focus.

Interviewer: *So this scale is not at all considered, to a very little extent, to some extent, to a great extent, to a very great extent.*

Interviewee: Let's go to some extent.

Question 32a

Interviewer: *Was there a formalised technique that you used for that?*

Interviewee: Bow tie was one.

Question 33

Interviewer: *Is Structural Health Monitoring planned, or currently in use on your equipment for use in Structural Integrity risk management?*

Interviewee: Yeh.

Question 34

Interviewer: *What methods are used to provide confidence in the risk analysis conducted, for example independent review, consensus of opinion, audit?*

Interviewee: Yeh I was going to say audit, audit and the review, consensus I'm not sure of, possibly but nothing I'm aware of.

Question 35

Interviewer: *Is assurance activity carried out to validate the effectiveness of your structural integrity risk analysis?*

Interviewee: Er yes.

Question 35a

Interviewer: *Who conducts it?*

Interviewee: ITEA, who was that *[other company name]*, oh no it wasn't *[other company name]*, it was *[other company name]*.

Interviewer: *So they're external to the organisation but within the same industry?*

Interviewee: Yeh.

Interviewer: *Is there any internal audit, or self audit.*

Interviewee: We have MAR, BMAR/MAR, from a historical document perspective, aging aircraft audits.

Interviewer: *So they're probably internal audit rather than self-audit?*

Interviewee: I think they're contractors but it's effectively internal, yeh.

Interviewer: *Do you ever see anyone outside of the organisation coming in? and outside of aviation, have you seen that?*

Interviewee: Not sure I have, top of my head.

Question 35b

Interviewer: *Do you believe the assurance activity is effective.*

Interviewee: Er yeh. I mean example being....if you want an example? So the life extension programme for the aircraft I worked on, *[aircraft type]*, the OEM did all the maths and analysis to say here's what we think we can extend it to, and the independent said actually we disagree with you, we're going to go to 60% of that, so me saying it's effective because it's perhaps more restrictive than the OEM.

Interviewer: *And they were actually looking at it as well?*

Interviewee: Well yeh. I don't think they disagreed with the maths, I think they wanted more evidence and it wasn't provided.

Question 36

Interviewer: *What changes might be made to improve your organisation's structural integrity risk management?*

Interviewee: Historical trending, if you like, what's the word, audit trail. A much much better way of managing audit trail of decisions and of changes, and how they interact. Great example being *[aircraft type]* where there was several dozens of three or four draw filing cabinets per aircraft with all sorts of different *[modifications]* and *[technical instructions]* and decisions about, yeh that crack this length, and all that sort of thing, and you wanted to sort of decide something new, look back on interactions with other elements from that cumulative risk type thing or historical trending, it's very very difficult to do.

Question 37

Interviewer: *Ok. And last question, Do you have any comments on the answers provided in this section?*

Interviewee: No I don't think so.

Interviewer: *Thank you very much, that is 15 minutes over, over the hour, sorry. It's deliberately intensive because I need that level of detail that we've gone to, specifically to garner the qualitative elements that I'll shortly be going through to pick out any key points across things and there's a lot of commonality that's starting to come out, so it's good to see.*

Interviewee: So I mean in terms of how I answered, it's actually quite difficult really, I felt unsure of how to answer a lot of those questions.

Interviewer: *Yeh I've had that as feedback. It's all part of me learning how to be a researcher, which is all part of the PhD process.*

Interviewee: I didn't mean it in that respect. Sometimes it was understanding the question, but even when I understood it, it was, oh that's kind of put me on the spot a little bit, it's given me a little bit of an eye opener, and well the *[aircraft structural integrity course]* was what 6 years ago I did that, probably the last formal training I had in structural health was the *[aircraft structural integrity course]* and maybe the RCM that I did, other than that it's just been on the job-type training.

Interviewer: *I'm just going to hit stop now, so if there's anything else you wanted recording...*

Interviewee: Nope.

Interviewee code	191220192100 (ID 61)
Organisation industry	Aviation.
Organisation category	Operator/end user
Role when handling risk	Risk analyst

Interviewer: *The date is 19 Dec 2019, are you happy for me to audio record this interview?*

Interviewee: I am.

Question 1

Interviewer: *Thank you very much. [Informed consent explanation and discussion]. So the first section is about setting the context and they're generalised, looking for your organisation and experience baseline. I'm going to make an assumption that your industry is 'Aviation'?*

Interviewee: Correct!

Question 2

Interviewer: *Which category best describes your organisation, are you within the operator/end user category, regulator/policy maker, or designer/equipment manufacturer, or other?*

Interviewee: I would say operator/end user.

Question 3

Interviewer: *What's your role within the organisation? These types of questions are probably, because of the role you're in now is less relevant now, but I've had some colleagues who've said they're an "aerosystems engineer" or an engineering manager, but I've equally had some people who've refer to the role that they were in when they were risk managing structure.*

Interviewee: Ok so, it's probably, best to go with the....I don't know, aerosystems engineer.

Question 4

Interviewer: *Ok. In your experience do organisation risk management policy specifically consider SI risk? Yes, no, or you're not aware of any organisational risk management policy.*

Interviewee: Yes, they do.

Question 5

Interviewer: *And within the context of equipment structural integrity, which of these best describes your role or the role most usually performed when handling risk? Are you a risk analyst, a risk decision-maker, a commentator or an advisor and I have the definitions if you want me to clarify any of those.*

Interviewee: I guess so my query, in terms of the survey, I would immediately say maybe a risk decision-maker but in our great duty holder construct that may not necessarily be accurate. So I'll go with risk decision-maker but it's basically assessing is there a risk there or not.

Interviewer: *So I've got that's an individual who decides upon the course of action for a given risk. They may make consideration given other factors outside of the particular risk envelope, e.g. operational, financial, commercial or other business constraints. Personally I think that's fine, but I've had individuals who've been [airworthiness delegation] holders, who've said they're a risk*

analyst or a risk decision-maker it doesn't really make too much of a difference.

Interviewee: What's a risk analyst?

Interviewer: *A Risk Analyst is an individual who produces a risk assessment based upon observations, beliefs and information available at the time of the analysis.*

Interviewee: Yeh.....go for risk analyst.

Question 6

Interviewer: *That's the end of section 1, section 2 addresses the analysis of in-service structural integrity risk. So can you describe the current process for conducting risk analysis for in-service structural integrity issues, from identification or realisation of the issue to the point when the risk is considered managed.*

Interviewee: Ok, so I will answer this drawing on experience from the *[aircraft type]* Project Team, probably more latterly in the time. I'd say the process would be: once the issue is identified you perform an immediate analysis which assesses the, almost the worst case outcome. Is it something that leads to a catastrophic failure and loss of the platform or will it be an aspect that just effects performance without any real safety impact. Once you've done that, you can process it into the correct kind of modelling to see if some of those risks are already captured within the FTLM *[fault tree loss model]*, which we used. If there wasn't and you know adjust the scoring of that accordingly give that new information. If there wasn't an existing hazard then you would seek to potentially create another one once you've done some further analysis on the risk itself and possible mitigations. I'd say if you could almost immediately implement some mitigations which reduce the risk to a tolerable level then you wouldn't formally issue a hazard for it and I guess expose that more widely to the organisation. It was only if you felt that with mitigations there would be a residual risk, or if you felt that the mitigations wouldn't completely reduce that risk to a tolerable level that you would then raise a separate hazard to that capture that which would affect the overall platform safety target. Now, the difficulty there, which will come to I'm sure, is in how to quantitatively express that increased risk, and so we would always take the view of there's an inherent design target of the aircraft with the structural integrity so we would try to make a judgement call on how much this particular risk would alter adherence to that design target, based on previous knowledge and experience. But then again, the purpose of your phd.... that is accetably very vague and quite likely hugely inaccurate.

Interviewer: *You mentioned about tolerable, if you're making a judgement whether something is tolerable or not, how did you decide on that? What's the process?*

Interviewee: So for me, again, so we've got a tolerable risk envelope for operating the aircraft, so if you had a hazard emerge, structural in nature, if through mitigations, you could confidently say there's no greater chance of that hazard being realised than there was before this information came to light, or this issue started, then I think you can address that as referring to tolerable in terms of risk analysis. So the example would be, if you have a fatigue crack and you have the data from the fatigue test specimen which tells you what the propagation is, you apply all the design safety standards with regards safety factors for inspection intervals etc. to bring that down to a point that you are confident that you carryout routine

inspections on that crack initiation area or propagation, in-line with those def stan documents that we work to, then you could be happy that you've returned it to a tolerable position because you've met something, which in accordance with defence standards meets the design target of $1x10^{-6}$ failure rate, so you've got to bring it back to a tolerable position. Now if you couldn't carry out those inspections at that interval because say you can only access them during a minor maintenance and it was very intrusive and you didn't want to do that so you were going to move it a less frequent inspection regime, then you're moving outside of that tolerability range and therefore that risk needs to be catergorized and accepted.

Interviewer: *So that tolerable envelope you said, was that a quantified envelope?*

Interviewee: I would say, yes, certainly within the *[aircraft project team]* we would work within a tolerable risk envelope which was the platform loss rate, so if what we were doing wouldn't affect the platform loss rate then that's fine. Obviously everything effects the platform loss rate in some way, so it's kind of tricky, so in terms of the structural [unclear] I was talking about it comes back to the...you know....inherently the aircraft should have been designed to have a catastrophic structural failure every 1 million flying hours, so as long as everything you're doing is adhering to meeting that 1 in a million, you're maintaining that tolerable region. That's an inherent design safety target, stemming from aircraft design and manufacture from many years ago, and I believe, accepting that I've been out of the *[regulatory and type airworthiness]* space for a little bit, I certainly believe that it used to be regulation that you had to achieve that $1x10^{-6}$ because I always remember there was some interesting discussion there that transport aircraft with 200 people had the same platform loss rate as a fast jet with ejection seats but hey there you go. So I believe that is still the case that that is still in the regulation set. Tolerability probably most of the guidance interms of what a tolerable risk is, is found from HSE literature in terms of first, second, and third party risks and what is deemed acceptable for an employer. I guess to answer your question, yes I feel it was defined at that time and we understood it. I think it varies between different platforms, do organisationally we do a good enough job of publicising what our tolerable risk appetite is. Probably not, but that's probably a function that we don't quantify everything perhaps so much. We tend to lean on ALARP, lets reduce the risk as much as we can...

Interviewer: *...which is largely a judgement-based factor.*

Interviewee: Yes. But I say that's just where we are culturally. I have found when I've worked with the US Air Force for example, they're a lot more quantitative with their assessment and they will have a much more rigid CBA process for things, so they kind of ear on the numbers rather than on an as low as reasonably practicable qualitative argument, so they will have a quantitative CBA versus a wordy answer.

Question 7

Interviewer: *Thank you. So I've got a few questions throughout which are a likert scale between strongly disagree and strongly agree, there is a neutral point and there is, so it goes Strongly Disagree, Disagree, Slightly Disagree, and then it's mirrored into the agree phase. So I'll give you warning before I read the phrase out. So*

the first one is: In my organisation, structural integrity risk analysis is an activity that is valued.

Interviewee: ... ooo....I'm pausing because I'm just wondering in comparison to other things is it given the weighting it perhaps deserves. [checks the options available] Agree.

Question 7a

Interviewer: *Why do you believe that?*

Interviewee: Because in my experience whenever there has been a SI risk which has been raised it's been taken seriously, there's always been funding available for structural upgrades that affect safety on platforms. I feel we have a robust *[airworthiness delegation]* process whereby we put people into positions, and there's an element of review there. It would maybe be nice to see some.. you know we have bespoke courses like the *[structural integrity course]*, maybe's there's something between a 1 years masters degree in aircraft design and 5 day course, I kind of feel there's probably something in the middle that feels like the right answer if we were doing this properly because again we shouldn't be in a position where your first 6 months on the job were kind of learning whereby structural integrity hazards are concerned, so that's why I haven't strongly agreed. But I think agree because where we do have risks identified they are taken seriously by the organisation.

Question 8

Interviewer: *Within your organisation, are structural integrity risks that arise in-service described numerically or non-numerically/verbally/descriptively? Or both?*

Interviewee: Both

Question 8a

Interviewer: *How would you rate the accuracy of your organisation's quantitative risk analysis of structural integrity? This is Very poor / Poor / Fair / Good / V.Good?*

Interviewee: It's somewhere between poor and fair. Er...Poor.....it's very dependent on what it is. I'll go for Fair.

Question 8b

Interviewer: *If you're using a situation which describes a risk verbally, then why in that case SI risk not quantified?*

Interviewee: Because it feels impossible to come up with a reference point to make an informed quantitative estimate.

Question 8c

Interviewer: *So this is another likert question. In my opinion, a quantitative/numerical value is necessary for the assessment and management of in-service structural integrity risk? What level of agree or disagree?*

Interviewee: I would disagree because there are I think for some examples I would agree, but in some cases you can offer a very effective mitigation with a reasonably low level of pain which means I think you can just mitigate that risk effectively without needing to go through painful analysis to quantify.

Question 8d

Interviewer: *So to follow up from that, In my opinion, a qualitative (verbal or descriptive) interpretation is necessary for the assessment and management of in-service structural integrity risk?*

Interviewee: Agree.

Question 8e

Interviewer: *Do you have any comments to support or expand those answers other than what you've mentioned already.*

Interviewee: So you need some qualitative assessment because that is a sort of mechanism of capturing where you got your number from.

Question 9

Interviewer: *When handling in-service structural integrity risks, are you required to assess structural integrity risk relative to a threshold level?*

Interviewee: Yes.

Question 9a

Interviewer: *What's the nature of that threshold level, for example loss rate per FH.*

Interviewee: So loss rate per FH against an inherent design safety target.

Question 9b

Interviewer: *In operational situations which are information poor, time critical, or resource constrained, I believe a threshold level is appropriate. Strongly disagree, disagree, etc.*

Interviewee: Strongly Agree.

Question 9c

Interviewer: *How is the threshold level set? Regulator / Statutory / Internal organisational limits / Personal limits / Designer/manufacturer limits / Other (specify)?*

Interviewee: So it's regulatory in terms of the *[aviation regulator]* platform loss rate, statutory in terms of the HSE guidance for tolerability of risk.

Question 9d

Interviewer: *What processes do you use in support of SI risk analysis or decision making? So if I read these out and you say yes or no after each one. What If analysis?* Yes.

Interviewer: *Check lists?*

Interviewee: Yes.

Interviewer: *Safety Audits?*

Interviewee: Yes.

Interviewer: *Task Audits?*

Interviewee: Yes, I haven't seen them but they're there.

Interviewer: *Bow Tie?*

Interviewee: Yes.

Interviewer: *PRAT?*

Interviewee: Proportional RA Technique, no I don't know what it is therefore no!.

Interviewer: *Risk matrices?*

Interviewee: No.

Interviewer: *Do you not?*

Interviewee: Is that just the frequency impact...oh yes....

Interviewer: *I know this is your questionnaire, but I'm pretty sure we did use them!*

Interviewee: Risk matrix, I thought you coming up with something more you know.

Interviewer: *Fancy, no!*

Interviewee: Is that a quantitative technique though, you come out with a number from a matrix from which it depends how you derived your quantitative input.

Interviewer: *Garbage in is garbage out at the end of the day! Quantitative Risk Assessment, QRA?*

Interviewee: No.

Interviewer: *Probability Distributions?*

Interviewee: I'm going to say yes to that because when you're doing the fatigue crack analysis, so if you had an incident on the *[aircraft fatigue test]* or something and you're looking at propagation rates etc. where you get your safety factors, it comes from doing your standard deviations and all that good stuff, so while we don't crank the numbers ourselves in the PT so much we were working with the rules of thumb that applied those techniques.

Interviewer: *Event Data Models, which are like historical data, frequency of failure?*

Interviewee: Yeh.

Interviewer: *These are hybrid approaches, Human Error Analysis Technique?*

Interviewee: No.

Interviewer: *Fault Tree Analysis?*

Interviewee: Yes.

Interviewer: *Event Tree Analysis?*

Interviewee: Yes.

Interviewer: *Risk Based Maintenance?*

Interviewee: There are some people that would say yes to that but I'm going to say no.

Interviewer: *Epistemic Models?*

Interviewee: No.

Interviewer: *Bayesian Networks?*

Interviewee: No.

Interviewer: *This grouping is stochastic processes, Monte Carlo simulation?*

Interviewee: No.

Interviewer: *Bayesian Analysis?*

Interviewee: No.

Interviewer: *Probabilistic distribution for key variables?*

Interviewee: I'm going to say yes to that for the reasons that I highlighted earlier for accepting probabilistic analysis.

Interviewer: *Bespoke numerical models for a given situation?*

Interviewee: Yes.

Interviewer: *Neural networks?*

Interviewee: No.

Interviewer: *Any others that I haven't mentioned?*

Interviewee: No I think you've covered them.

Question 9d1

Interviewer: *Do you believe these processes are appropriate for structural integrity risk analysis in your organisation?*

Interviewee: Yes I believe they're appropriate, I wouldn't say that I don't think they could be improved on.

Question 9d1a

Interviewer: *Why's that?*

Interviewee: Well I feel we're using probabilistic analysis, we could do more to accurately model our quantification of those risks, I don't know what the answer is to do that. But there's a lot of clever statistics out there.

*[*Discussion about the direction of the study]

Question 10

Interviewer: *Is your organisation's risk analysis methodology proactive, enabling anticipation of risk from in-use changes or modifications to usage/configuration, or is it reactive, simply responding to issues that occur?*

Interviewee: I would say it is predominantly reactive, but it is proactive because of things like the Statement of Operating Intent and Usage etc. we do try to forward predict issues with mixed success across the MOD I might add, but I would say yes with room for improvement in that area.

Question 10a

Interviewer: *Do you use any specific methods to identify risks in that guise if you're proactive?*

Interviewee: I think the main thing is to do the calendar based reviews against design assumptions against actual operational usage, similarly with the maintenance the R&M process looks at failures and if there are spikes it will try and do some more analysis to try and understand what might be happening, accepting that's a bit reactive when you start having failures. So I guess actually the main thing is how we use the aircraft and we try and anticipate if that would have any impact or lead to any additional failures which I would define as proactive, as I say room for improvement in that though.

Question 11

Interviewer: *The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'likelihood' of occurrence.*

Interviewee: Disagree, just because of the accuracy of that likelihood.

Question 12

Interviewer: *The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'consequence' of occurrence.*

Interviewee: Strongly Agree.

Question 13

Interviewer: *More often than not, there are uncertainties within my analyses of structural integrity risks So uncertainties here are information or a value that is not known accurately.*

Interviewee: Strongly agree.

Question 13a

Interviewer: *Are uncertainties communicated in risk analyses regarding operational structural integrity?*

Interviewee: Yes.

Question 14

Interviewer: *More often than not, there are assumptions within my analyses of structural integrity risks.*

Interviewee: Strongly Agree

Question 14a

Interviewer: *Are assumptions communicated in risk analyses regarding operational structural integrity?*

Interviewee: Yes.

Question 15

Interviewer: *More often than not, I handle structural integrity risks conservatively. Conservatism here is cautionary thinking, over-estimation of the risk resulting in additional measures or values added to provide a safety factor given uncertain quantitative outcomes.*

Interviewee: Strongly Agree

Question 15a

Interviewer: *When conservatism is applied, is it communicated in risk analyses regarding operational structural integrity?*

Interviewee: Yes.

Question 16

Interviewer: *To what extent do you believe human subjective judgement (from expert opinion or belief) influences the analysis of in-service structural integrity risks? This is a scale from Not at all considered / To a very little extent / To some extent / To a great extent / To a very great extent.*

Interviewee: To a very great extent.

Question 16a

Interviewer: *Just to clarify, a human subjective judgement is a judgement influenced by human opinions, feelings or tastes. Are human subjective judgements communicated?*

Interviewee: No. That's interesting.

Question 17

Interviewer: *Likert scale - I am confident in the structural integrity risk analysis method applied for handling operational risks.*

Interviewee: Agree.

Question 17a

Interviewer: *Why?*

Interviewee: Because I feel the method I would say the only answer is conservative so it allows us to not over react but I feel we're always on the side of caution which offsets the difficulty we have in quantitatively applying, yeh, sorry, accurately applying a quantitative assessment to it. The point you bring up on the judgements being subjective is also valid, but then I feel that whilst we don't communicate that subjectivity, all of our subjective judgements are subject to peer review both within the design organisations and within the *[aircraft airworthiness authority]* and then indeed both organisations check each other, so it's a collection of view points. I guess the danger there would need to be organisationally within the two organisations or indeed both if there was a way of thinking developed. There's probably a moment there where you talk about subjective viewpoints to make sure that you have a good diversity of thought in those teams which speaks to both having experienced individuals and also rotating people around from other areas of SI with similar experience but might not have come from a team with the same cultural values, to avoid the group think moments.

Interviewer: *Or bias*

Interviewee: Yeh.

Question 18

Interviewer: *Do you have any other comments on that section that we've just gone through?*

Interviewee: No, I don't think so.

Question 19

Interviewer: *Section 3 is all about decision making specifics of structural integrity risks. What information is essential in order to make a decision upon an in-service structural integrity risk?*

Interviewee: The What-If consequences of failure, and whatever you require to make a subjective judgement on the likelihood.

Question 20

Interviewer: *Describe the decision-making process regarding in-service structural integrity risks.*

Interviewee: How's that different? Ah you asked how recorded before. Conscious I'm not repeating my answer from earlier overly, in terms of DM, I'll focus on the first element of that, which is an issue is highlighted, you immediately ,I would say, do the what if analysis because if the what if analysis says you know if this piece of tertiary structure cracks then is easily noticed and has no real impact be it a piece of plastic in the cockpit or something, which has no function other than a cowling, and wouldn't become a loose article hazard, then you would be fairly relaxed. Conversely if it was a main wing spar you wouldn't be. So you do the what if analysis look at that and do some second order things, for example I mentioned loose article etc. earlier, and once you've got a good what the what if analysis is, you look at what your perceived likelihood of that is and once you've looked at the worst case you then look at what's the most likely to happen and what the perceived likelihood of that is. With those two pieces of information in conjunction with a risk matrix you can then start to form a view of the overall severity of the risk and what steps would be appropriate to mitigate it.

Question 21

Interviewer: *Do you ever use judgement when making or supporting a decision regarding in-service structural integrity risks?*

Interviewee: Yes.

Question 21a

Interviewer: *In which situations would you use judgement? When there is a lack of available expertise?* Yes.

Interviewer: *When time constrained?*

Interviewee: Er...yes.

Interviewer: *When I disagree with advice?*

Interviewee: Yes.

Interviewer: *When there is insufficient resource to support a decision choice that would otherwise be available?*

Interviewee: Yes.

Interviewer: *When there is insufficient information available to support an objective analysis?*

Interviewee: Yes.

Interviewer: *When I believe that a particular decision choice is 'good enough'?*

Interviewee: Yes.

Interviewer: *When I believe that technical information is conservative?*

Interviewee: Yes.

Interviewer: *Any other times?*

Interviewee: Just those!

Interviewer: *But genuinely, are there any others?*

Interviewee: I think to be honest you use judgement in almost all steps of it, because just because the information isn't there, you then are presented with an option of 'stop what you're doing' but you would still need to justify that, so you're still then using your judgement. Because the default position of the organisation is to continue as normal you know the aircraft are flying in the same way if it was checking the structural integrity of a bridge cars are still driving over it, if you didn't have information and you wanted to stop the aircraft flying or stop cars driving over the bridge, you probably wouldn't be able to do that just simply on the basis of not having the information available yet, you would need to make a judgement on the severity. If there was a really big crack in the bridge and you didn't understand it, yep, you would stop people crossing the bridge. If there was a really big crack in an Structurally Significant Item and you couldn't understand where it had come from or why, you would absolutely consider a pause in flying particularly if you saw it on other occasions, while you understood more, but there you're using your judgement based on what your perceived severity is again. If it was something minor on the bridge that was maybe a decorative functioning piece of brickwork had just fallen off, or similarly a piece of tertiary structure or something on the aircraft you would not view that as an issue. But, so it's completely circumstantial really, so you have to use your judgement I think.

Question 21b

Interviewer: *When you use judgement, is your judgement holistic (by gut, by intuition, or by instinct), or do you decompose your judgement upon the situation into constituent parts?*

Interviewee: I try to decompose it into constituent parts.

Question 21b1

Interviewer: *Are the judgements you make numerical or non-numerical/verbal?*

Interviewee: Non-numerical, it has been both but it is predominantly non-numerical.

Question 21b2

Interviewer: *When you decompose the situation, how do you represent the decomposed form?*

Interviewee: I think I break it down into a similar process as to what if, you kind of follow a point through in terms of what that item does, what the implications of failure are for that and surrounding structure, and then that and the surrounding systems in terms of loose articles etc. You then take a look at, when you're talking about probability you then break it down to understand what probability you've seen so far when it occurs, what the status of the rest of the fleet is, because you know is it something that could've been picked up on routine inspections. So you kind of follow through those processes to try and understand, put some thinking behind those two things, the impact and probability.

Interviewer: *So you've got micro processes almost for each element of your decomposed situation.*

Interviewee: Yeh.

Question 21c

Interviewer: *What do you find difficult when using judgement to support decisions regarding in-service structural integrity risk?*

Interviewee: A lack of experience to draw on, specifically regarding structural integrity.

Question 21d

Interviewer: *How could you be supported better when using judgement to support decisions regarding structural integrity risk?*

Interviewee: I think we could be more exposed to real life SI scenarios and situations so that you get an understanding from a broader base of applications.

Question 21e

Interviewer: *Have you received training in applying judgement to support decisions regarding structural integrity risk?*

Interviewee: No. Well. Hmmm..a bit unfair actually. I'll say no because it's not really a scientific 'here's how you use your judgement', it's generic RA training.

Question 22

Interviewer: *In operational situations, decisions regarding a structural integrity risk are clearly justified. An Operational situation is a situation that is time critical, as well as information poor and/or resource constrained.*

Interviewee: Yeh I'd agree with that.

Question 22a

Interviewer: *Is there a mechanism for ensuring that structural integrity risk analyses are justified?*

Interviewee: I guess that mechanism is, the only thing is peer review, which isn't compulsory, or you have by virtue of being an *[airworthiness delegation]* holder you have to authorise it. Not sure if that classifies as suitably robust from a systems perspective, so I'll say No.

Question 23

Interviewer: *Are you more comfortable with numerical or non-numerical/verbal/descriptive representations of risky decisions regarding in-service structural integrity?*

Interviewee: Clearly I'm more comfortable with numerical provided they're accurate.

Interviewer: *So numerical?*

Interviewee: Yeh, if they're available but that is accurate numerical figures for example based on fatigue test specimens, not wild ass guesses.

Interviewer: *Engineering judgement you mean?!*

Interviewee: Exactly!

Question 24

Interviewer: *Within your organisation, are decisions upon structural integrity risk assessments "risk-informed" or "risk-based"? Risk-informed is Risk assessment results support decision making but are not the only consideration, and Risk-based is Risk assessment results form the sole basis for decision making.*

Interviewee: Not sure what to say for that one. Give me the question again. (repeated). Risk-based but you could argue a split between the two.

Question 25

Interviewer: *In operational situations, have you received pressure or influencing force to make a decision on a structural integrity risk?*

Interviewee: Yes.

Question 26

Interviewer: *In your opinion, in-service structural integrity risk analyses are used effectively by decision-makers. Strongly disagree to strongly agree?*

Interviewee: Agree.

Question 26a

Interviewer: *Why?*

Interviewee: Because they're reviewed and considered and largely the views of the experts who are making the subjective judgements are taken as accurate and acted upon.

Question 27

Interviewer: *Considering situations that you have seen of structural integrity risk management decision making, do you believe decision-makers understand the assumptions and uncertainties presented to them from structural integrity risk analyses? This is broken into a kind of part A and part B. So uncertainties is taken first, which is between Never / Sometimes / About half the time / Most of the time / Always / Unsure.*

Interviewee: Most of the time.

Interviewer: *And Assumptions?*

Interviewee: About half the time.

Question 27a

Interviewer: *Are you able to list any common reasons that decision-makers do not understand assumptions with structural integrity risk?*

Interviewee: Lack of familiarity with probabilistic methods and how they're derived for inherent design safety targets with respect to fatigue; and just wider structural design assumptions.

Question 27b

Interviewer: *And any common reasons that decision-makers do not understand uncertainties with structural integrity risk.*

Interviewee: Similar reason , I think that sometimes people get fixated on the number and as soon as they see the number they assume it's accurate so even when you put uncertainty to it, it's not quite absorbed.

Question 28

Interviewer: *Are there any comments on the section on decision making.*

Interviewee: No.

Question 29

Interviewer: *Last section which is considering the management of structural integrity risks arising in-service, and the handling of the risk analysis conducted. Were you previously, or have you currently been formally delegated the authority to own and decide upon structural integrity risks in your position?*

Interviewee: Yes.

Question 29a

Interviewer: *Did/do you own this risk solely or in conjunction with others?*

Interviewee: I'm going to say in conjunction with others and it comes back to my earlier point when you asked whether I was a risk decision-maker or risk analyst, and I'm conscious that I've just answered that I make decisions. But I would say that substantiated risks to life on the platform were transferred to, at the very least, *[senior aircraft airworthiness authority]*/duty holder level but at the same time I think the role I had in dealing with the lower levels of risk and mitigation did lead to decisions, thus arguably mixed responses and inconsistency across....

Question 29b

Interviewer: *No I'm picking up the semantics of risk in Structural Integrity is not easily captured in a survey. The reason I'm asking these questions is to form an idea for why, and how, and to what level individuals have handled SI risk, and whether they had an opportunity to pass risk up or whether they were left solely to it "right you're the structures man it's up to you", that's the kind of basis for that thinking. If you were/are not able to decide upon a particular in-service structural integrity risk, who does make the ultimate decision?*

Interviewee: The *airworthiness authority* for the platform.

Question 30

Interviewer: *Your organisation's risk management method enables you to maintain a transparent audit trail, which details the history of the treatment and ownership of in-service structural integrity risks. Strongly disagree to strongly agree?*

Interviewee: Agree.

Question 30a

Interviewer: *What measures/aspects of the risk management framework enable the process to be considered as transparent?*

Interviewee: It was all hosted on a shared environment, be that on *digital maintenance manual* for queries, the Fault Tree Loss Model on sharepoint, or just the structures archive.

Question 31

Interviewer: *Who do you believe acts upon the risk analysis conducted for in-service structural integrity issues?*

Interviewee: (Operator/end user) Well yes, I suppose if we identified a big issue they would act on it, the example being when we grounded an aircraft when we had issues with fatigue in the wings.(Regulators/internal dept.) No, they don't really act on it. (Manufacturer) Yes (equipment owner) Yes (line manager) Yes (Another department internal to organisation) No I don't think so. (Any others?)

Question 32

Interviewer: *To what extent does your structural integrity risk management strategy explicitly try to identify rare events, 'black swans' or 'unknown unknowns'? So a black swan is a surprising event that may not have been considered when the decision was being made, but that would have changed the decision if it had been considered. And this is a scale between Not at all considered / To a very little extent / To some extent / To a great extent / To a very great extent.*

Interviewee: It's either not at all or to a very little extent. Just trying to think if I can think of some examples where...I think we kind of always set, it comes back to the teardown and part of the reason for that, so

there was activity taken. The whole point of that was to validate everything, so you can't say actually based on that that we didn't do much, so can I go for the middle ground on that one...to some extent.

Question 32a

Interviewer: *Do you use a formalised technique to identify rare events?*

Interviewee: No I think is all I can say is it's that structural teardown and sampling where we've done all the analysis, we've got some in-service surveying but we still have a detailed look in an effort to identify unknown unknowns.

Interviewer: *Which was a benefit of having older aircraft, rather than, a new aircraft might not have that luxury.*

Interviewee: And the reasons I didn't give a stronger answer for that is that I don't think we undertook that on a theoretical basis, it was based on the age of the platform we'll go and do some sampling rather than earlier in the life cycle.

Question 33

Interviewer: *Is Structural Health Monitoring planned, or currently in use on your equipment for use in Structural Integrity risk management?*

Interviewee: Ah well....on *[aircraft type]*...no.

Question 34

Interviewer: *What methods are used to provide confidence in the risk analysis conducted, for example independent review, consensus of opinion, audit?*

Interviewee: All of those, independent review, peer review, audit, I guess covered in independent review, that's both independent internally and obviously we used independent safety advsisors as well, so that's probably the main one.

Question 35

Interviewer: *Is assurance activity carried out to validate the effectiveness of your structural integrity risk analysis?*

Interviewee: I'm going to say no because I don't feel that anyone who, hmmm... sorry can you give me the question again (question read again). No because it's not formal assurance, other than the independent safety advisor, when people came in to so called assure us, they were looking at process compliance rather than reviewing our actual risk analysis, the decision we came to. So we could've been talking complete cobblers but as long as we'd followed the process it would've been fine, so I'm going to say no.

Question 36

Interviewer: *What changes might be made to improve your organisation's structural integrity risk management?*

Interviewee: I think it is a better grasp of possible models to use for numerical analysis. I think it is a greater kind of assurance from SQEP individuals able to review the decisions and I think it's greater sharing of structural integrity incidents and experiences across the organisation.

Question 37

Interviewer: *Have you got any comments on the answers provided in this section?*

Interviewee: No thank you.

Interviewer: *And that is the end of the questions. Thank you very much.*

Appendix D

Main survey questions

The following pages contain the questions, answer options and glossary definitions provided to participants for the Phase 2 primary data collection survey, described in Section 3.

In this first section, generalised questions outline the context of respondent organisations and experience of structural integrity risk.

1. Within which industry is your organisation? [List e.g. Aviation / Civil / Automotive / Nuclear / Rail / Other (specify)]

2. Which category best describes your organisation? [Operator/end user / Regulator/policy maker / Designer/equipment manufacturer / other (describe)]

3. What is your role within the organisation? [Open answer (max characters)]

4. In your experience, do organisational risk management policies specifically consider structural integrity risk? [Yes / No / No organisation risk management policy]

5. Within the context of equipment structural integrity, which of these best describes your role (or the role most usually perform) when handling risk? [Risk analyst / Risk Decision Maker / Commentator / Advisor]
Advisor *– an individual who may take a view and provide advice to a risk analyst or risk decision maker, but is not responsible for the decisions made within the risk management framework.*
Commentator *– an individual who may take a view on a given risk but has no particular responsibilities within the risk management framework.*
Risk Analyst *– an individual who produces a risk assessment based upon observations, beliefs and information available at the time of the analysis.*
Risk Decision Maker *– an individual who decides upon the course of action for a given risk. They may make consideration given other factors outside of the particular risk envelope, e.g. operational, financial, commercial or other business constraints.*

Section Two addresses the analysis aspects of in-service structural integrity risk.

6. Describe the current process for conducting risk analysis for in-service structural integrity issues, from identification/realisation of an issue to the point when the risk is considered managed. [Open answer]

7. In my organisation, structural integrity risk analysis is an activity that is valued [Strongly disagree / Disagree / Slightly Disagree / Neutral / Slightly Agree / Agree / Strongly Agree]

    a. *[All]* Why do you believe this to be the case? [Open answer]

8. Within your organisation, are structural integrity risks that arise in-service described numerically or non-numerically/verbally/descriptively? [Numerically / Verbally / Both]

    a. *[Numerically / Both]* How would you rate the accuracy of your organisation's quantitative risk analysis of structural integrity? [Very poor / Poor / Fair / Good / V.Good]

    b. *[Verbally]* Why is structural integrity risk not quantified? [Open answer]

    c. *[All]* In my opinion, a quantitative (numerical) value is necessary for the assessment and management of in-service structural integrity risk? [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]

    d. *[All]* In my opinion, a qualitative (verbal/descriptive) interpretation is necessary for the assessment and management of in-service structural integrity risk? [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]

    e. *[All]* Do you have any comments to support or expand upon these answers?

9. When handling in-service structural integrity risks, are you required to assess structural integrity risk relative to a threshold level? [Yes / No]

a. *[Yes]* What is the nature of the threshold level? For example, your threshold level might be communicated as a loss rate per flying hour. [Open answer]

b. *[Yes]* In operational situations which are information poor, time critical, or resource constrained, I believe a threshold level is appropriate. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]

Operational situation *– a situation that is time critical, as well as information poor and/or resource constrained.*

c. *[Yes]* How is this threshold level set? [Regulator / Statutory / Internal organisational limits / Personal limits / Designer/manufacturer limits / Other (specify)]

d. *[Yes]* What processes do you use in support of SI risk analysis or decision making (tick all that apply)? [None / (Deterministic methods: (qualitative: What If analysis / Check lists / Safety Audits / Task Audits), (quantitative: Bow Tie, Proportional Risk Assessment Technique, Risk Matrix, Quantitative Risk Assessment), (Classical statistical approach: probability distributions / event data models (historical data,frequency of failure), (Hybrid: Human Error Analysis Technique / fault tree analysis / event tree analysis / Risk Based Maintenance / Epistemic models / Bayesian Networks), (stochastic processes: monte carlo simulation / Bayesian analysis / probabilistic distributions for key variables / bespoke numerical models for a given situation / neural networks), Other (specify)]

(1) *[All]* Do you believe these processes are appropriate for structural integrity risk analysis in your organisation? [Yes/No]

(a) *[Yes/No]* Why do you believe this to be the case?

e. *[No]* What measure/test is used to demonstrate a safe operating level in your organisation? [None / Open answer]

10. Is your organisation's risk analysis methodology *proactive,* enabling anticipation of risk from in-use changes or modifications to usage/configuration, or is it *reactive*, simply responding to issues that occur? [Proactive / Reactive] (Add comments)

a. *[Proactive/Reactive]* What methods do you use to identify risks?

11. The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'likelihood' of occurrence. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]

12. The method I employ for conducting analyses of in-service structural integrity risks generates a clear 'consequence' of occurrence. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]

13. More often than not, there are uncertainties within my analyses of structural integrity risks. (See glossary) [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]

Uncertainties *– information or a value that is not known accurately.*

a. *[Neither / Slightly Agree / Agree / Strongly Agree]* Are uncertainties communicated in risk analyses regarding operational structural integrity? [Yes / No]

14. More often than not, there are assumptions within my analyses of structural integrity risks. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]
Assumptions *– an activity or value that is presumed to be true, for the purposes of analysis.*

a. *[Neither / Slightly Agree / Agree / Strongly Agree]* Are assumptions communicated in risk analyses regarding operational structural integrity? [Yes / No]

15. More often than not, I handle structural integrity risks conservatively. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]
Conservatism *– cautionary thinking, over-estimation of the risk resulting in additional measures or values added to provide a safety factor given uncertain quantitative outcomes.*

a. *[Neither / Slightly Agree / Agree / Strongly Agree]* When conservatism is applied, is it communicated in risk analyses regarding operational structural integrity? [Yes / No]

16. To what extent do you believe human subjective judgement (from expert opinion or belief) influences the analysis of in-service structural integrity risks? [Not at all considered / To a very little extent / To some extent / To a great extent / To a very great extent]
Human subjective judgement *– a judgement influenced by human opinions, feelings or tastes.*

a. *[All]* Are human subjective judgements communicated in risk analyses regarding operational structural integrity? [Yes / No]

17. I am confident in the structural integrity risk analysis method applied for handling operational risks. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]

a. *[All]* Why do you believe this to be the case?

18. Do you have any comments on the answers provided in this section? [Open answer]

Section Three considers decision making specifics of structural integrity risks.

19. What information is essential in order to make a decision upon an in-service structural integrity risk? [Open answer]

20. Describe the decision-making process regarding in-service structural integrity risks. [Open answer]

21. Do you ever use judgement when making or supporting a decision regarding in-service structural integrity risks? [Yes / No]

a. [Yes] In which types of situation do you use judgement? (select all that apply) [When there is a lack of available expertise / When time constrained / When I disagree with advice / When there is insufficient resource to support a decision choice that would otherwise be available / When there is insufficient information available to support an objective analysis / When I believe that a particular decision choice is ‘good enough’ / When I believe that technical information is conservative / Other (specify)]

b. [Yes] Is your judgement holistic (by gut, by intuition, or by instinct), or do you decompose your judgement upon the situation into constituent parts? [Holistic / Decomposed]

(1) [All] Are the judgements you make numerical or non-numerical/verbal? [Numerical / Non-numerical/verbal / Both]

(2) [Decomposed] How do you represent the decomposed the situation? [Open answer]

c. [Yes] What do you find difficult when using judgement to support decisions regarding in-service structural integrity risk? [Open answer]

d. [Yes] How could you be supported better when using judgement to support decisions regarding structural integrity risk? [Open answer]

e. [Yes] Have you received training in applying judgement to support decisions regarding structural integrity risk? [Yes / No]

22. In operational situations, decisions regarding a structural integrity risk are clearly justified. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]
Operational situation *– a situation that is time critical, as well as information poor and/or resource constrained.*

a. *[All]* Is there a mechanism for ensuring that structural integrity risk analyses are justified? [Yes / No]

(1) *[Yes]* What is the mechanism that you use for ensuring that structural integrity risks analyses are justified? [Open answer]

23. Are you more comfortable with numerical or non-numerical/verbal/descriptive representations of risky decisions regarding in-service structural integrity? [Numerical / Non-numerical/verbal]

24. Within your organisation, are decisions upon structural integrity risk assessments “risk-informed” or “risk-based”? [Risk-informed / Risk-based]
Risk-informed *– Risk assessment results support decision making but are not the only consideration.*
Risk-based *– Risk assessment results form the sole basis for decision making.*

25. In operational situations, have you received pressure or influencing force to make a decision on a structural integrity risk? [Yes / No]
Operational situation *– a situation that is time critical, as well as information poor and/or resource constrained.*

26. In your opinion, in-service structural integrity risk analyses are used effectively by Decision Makers. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]
Risk Decision Maker *– an individual who decides upon the course of action for a given risk. They may make consideration given other factors outside of the particular risk envelope, e.g. operational, financial, commercial or other business constraints.*

a. *[All]* Why do you believe this to be the case? [Open answer]

27. Considering situations that you have seen of structural integrity risk management decision making, do you believe Decision Makers understand the assumptions and uncertainties presented to them from structural integrity risk analyses?
Risk Decision Maker *– an individual who decides upon the course of action for a given risk. They may make consideration given other factors outside of the particular risk envelope, e.g. operational, financial, commercial or other business constraints.*
Assumptions *– an activity or value that is presumed to be true, for the purposes of analysis.*
Uncertainties *– information or a value that is not known accurately.*

Uncertainties [Never / Sometimes / About half the time / Most of the time / Always / Unsure]
Assumptions [Never / Sometimes / About half the time / Most of the time / Always / Unsure]

a. *[Never / Sometimes / About half the time / Most of the time]* List upto 3 most common reasons that Decision Makers do not understand assumptions with structural integrity risk analysis? [Open answer]

b. *[Never / Sometimes / About half the time / Most of the time]* List upto 3 most common reasons that Decision Makers do not understand uncertainties within structural integrity risk analysis? [Open answer]

28. *[All]* Are there any comments particular to this section?

Section Four considers the management of structural integrity risks arising in-service, and the handling of the risk analysis conducted.

29. Were you previously, or have you currently been formally delegated the authority to own and decide upon structural integrity risks in your position? [No / Yes]

a. *[Yes]* Did/do you own this risk solely or in conjunction with others? [Sole owner / In conjunction with others]

b. *[Yes]* If you were/are not able to decide upon a particular in-service structural integrity risk, who does make the ultimate decision? [Open answer]

c. *[No]* Who is the ultimate Decision Maker regarding in-service structural integrity risks? [Open answer] ***Risk Decision Maker*** *– an individual who decides upon the course of action for a given risk. They may make consideration given other factors outside of the particular risk envelope, e.g. operational, financial, commercial or other business constraints.*

30. Your organisation's risk management method enables you to maintain a transparent audit trail, which details the history of the treatment and ownership of in-service structural integrity risks. [Strongly disagree / Disagree / Slightly Disagree / Neither / Slightly Agree / Agree / Strongly Agree]

a. *[Neither / Slightly Agree / Agree / Strongly Agree]* What measures/aspects of the risk management framework enable the process to be considered as transparent? [open answer]

31. Who do you believe acts upon the risk analysis conducted for in-service structural integrity issues? (select all that apply) [Operator/end user / regulator / manufacturer / equipment owner / line manager / another department internal to organisation / other (specify)]

32. To what extent does your structural integrity risk management strategy explicitly try to identify rare events, 'black swans'[1] or 'unknown unknowns'? [Not at all considered / To a very little extent / To some extent / To a great extent / To a very great extent]
Black swan *– a surprising event that may not have been considered when the decision was being made, but that would have changed the decision if it had been considered.*

a. *[To a very little extent / To some extent / To a great extent / To a very great extent]* Do you use a formalised technique to identify rare events? [Yes (specify) / No]

33. Is Structural Health Monitoring planned, or currently in use on your equipment for use in Structural Integrity risk management? [Yes / No / Unsure]

[1] Cox, L. A. (2012). Confronting Deep Uncertainties in Risk Analysis. *Risk Analysis*, *32*(10), 1607–1629. https://doi.org/10.1111/j.1539-6924.2012.01792.x

34. What methods are used to provide confidence in the risk analysis conducted (for example independent review, consensus of opinion, audit)? [Open answer/use interviews/pilot answers to form multiple choice]

35. Is assurance activity carried out to validate the effectiveness of your structural integrity risk analysis? [Yes / No / Unsure]

a. *[Yes]* Who conducts assurance activity? [self audit / internal audit (internal to organisation) / external audit (external to organisation, but within same industry) / external audit (external to organisation, but outside of your industry)]

b. *[Yes]* Do you believe the assurance activity is effective? [Yes / No]

36. What changes might be made to improve your organisation's structural integrity risk management? [Open answer]

37. Do you have any comments on the answers provided in this section? [Open answer]

END OF QUESTIONS

Questionnaire- specific glossary

- Advisor – an individual who may take a view and provide advice to a risk analyst or risk decision maker, but is not responsible for the decisions made within the risk management framework.
- Assumptions – an activity or value that is presumed to be true, for the purposes of analysis.
- Black swan – a surprising event that may not have been considered when the decision was being made, but that would have changed the decision if it had been considered.
- Commentator – an individual who may take a view on a given risk but has no particular responsibilities within the risk management framework.
- Conservatism – cautionary thinking, over-estimation of the risk resulting in additional measures or values added to provide a safety factor given uncertain quantitative outcomes.
- Negative risk / threats – uncertain potential outcomes that, if they occur, may have a detrimental effect for the organisation.
- Operational situation – a situation that is time critical, as well as information poor and/or resource constrained.
- Positive risk / opportunities – uncertain potential outcomes that, if they occur, may have a beneficial effect for the organisation.
- Risk Analyst – an individual who produces a risk assessment based upon observations, beliefs and information available at the time of the analysis.
- Risk Decision Maker – an individual who decides upon the course of action for a given risk. They may make consideration given other factors outside of the particular risk envelope, e.g. operational, financial, commercial or other business constraints.
- Uncertainties – information or a value that is not known accurately.
- Risk-informed – Risk assessment results support decision making but are not the only consideration.
- Risk-based – Risk assessment results form the sole basis for decision making.
- Human subjective judgement – a judgement influenced by human opinions, feelings or tastes.

Appendix E

Main survey screenshots



Supporting engineering judgement-based decision making to recover aircraft structural integrity risk under operational conditions.

Lead Researcher - Rich Green MEng(Hons) BSc(Hons) CEng MIET RAF

Foreword

"Accurate, transparent and timely decisions on risky events is crucial to assuring equipment availability and safety. Despite inherent safety factors and designed defence against equipment structural failure, unanticipated cracking and deterioration still occur. Load bearing and support structure useful life are threatened by variability in through-life factors such as environmental exposure, usage and equipment requirement changes, and other material deterioration threats. The risk management of such variabilities in structure is challenging, resource intensive, usually conservative and believed to be inconsistent across technical industries. The study pursued by Squadron Leader Green aims to broadly benchmark structural integrity (SI) risk management techniques and optimise current academic progress in SI risk analysis with real-world application. As an industry regulator, this research is essential to improving transparency and recognisable risk advice to decision makers charged with responsibility for complex decisions regarding SI, and further, seeks a systematic method from which risk analysts and decision makers can operate when faced with uncertain structural outcomes, delivering a consistent and justifiable outcome."
Wing Commander Paul Regan - Military Aviation Authority Certification - Structures & Aircraft Design Standards (S&ADS)

Informed Consent Outline

The study aims to progress practical aspects of risk management for structural integrity of equipment or infrastructure, dependent upon your most applicable background experience regarding structural integrity risk, as indicated in the foreword of Wing Commander Regan.

Whilst there is no immediate benefit to each individual in taking part in this research, the

Figure E.1: Practitioner survey screen capture no. 1.

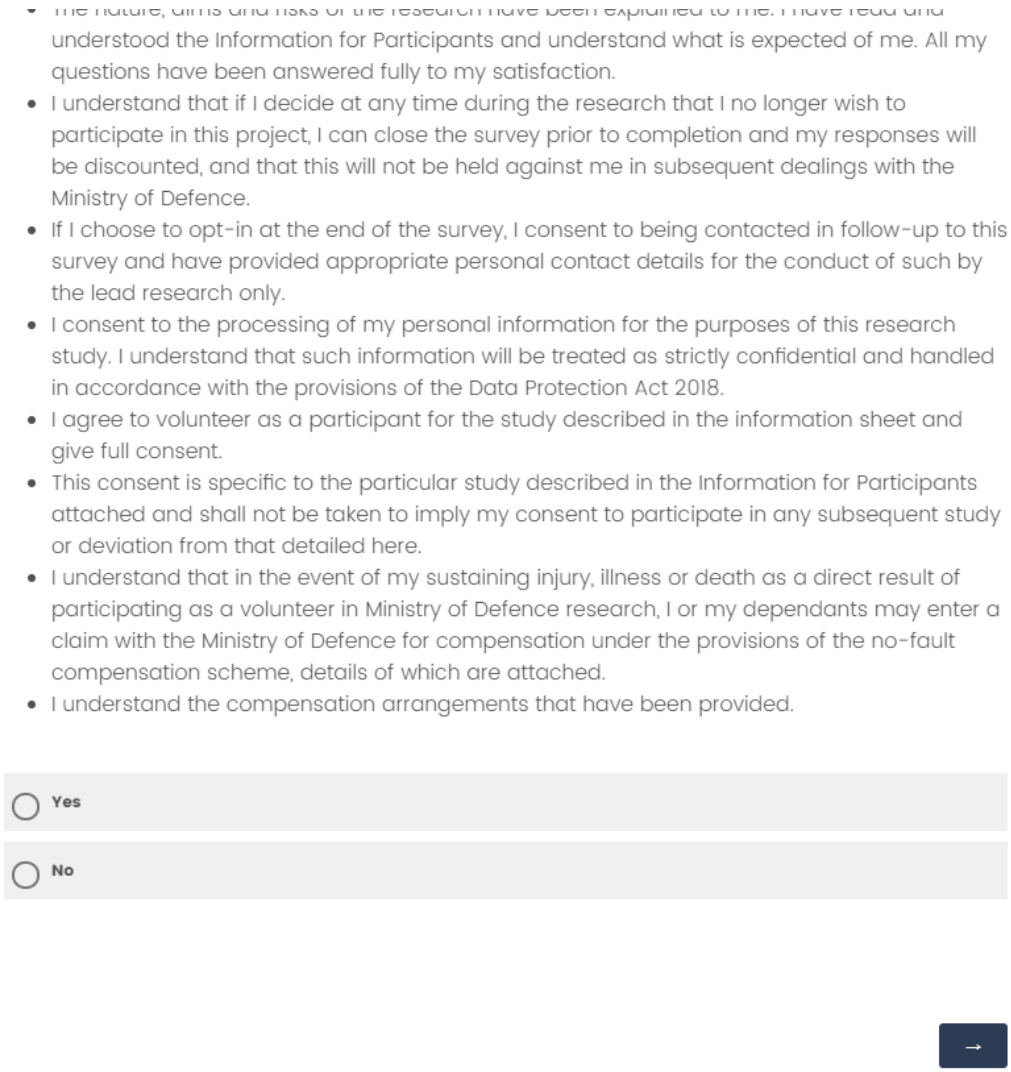


Figure E.2: Practitioner survey screen capture no. 2.

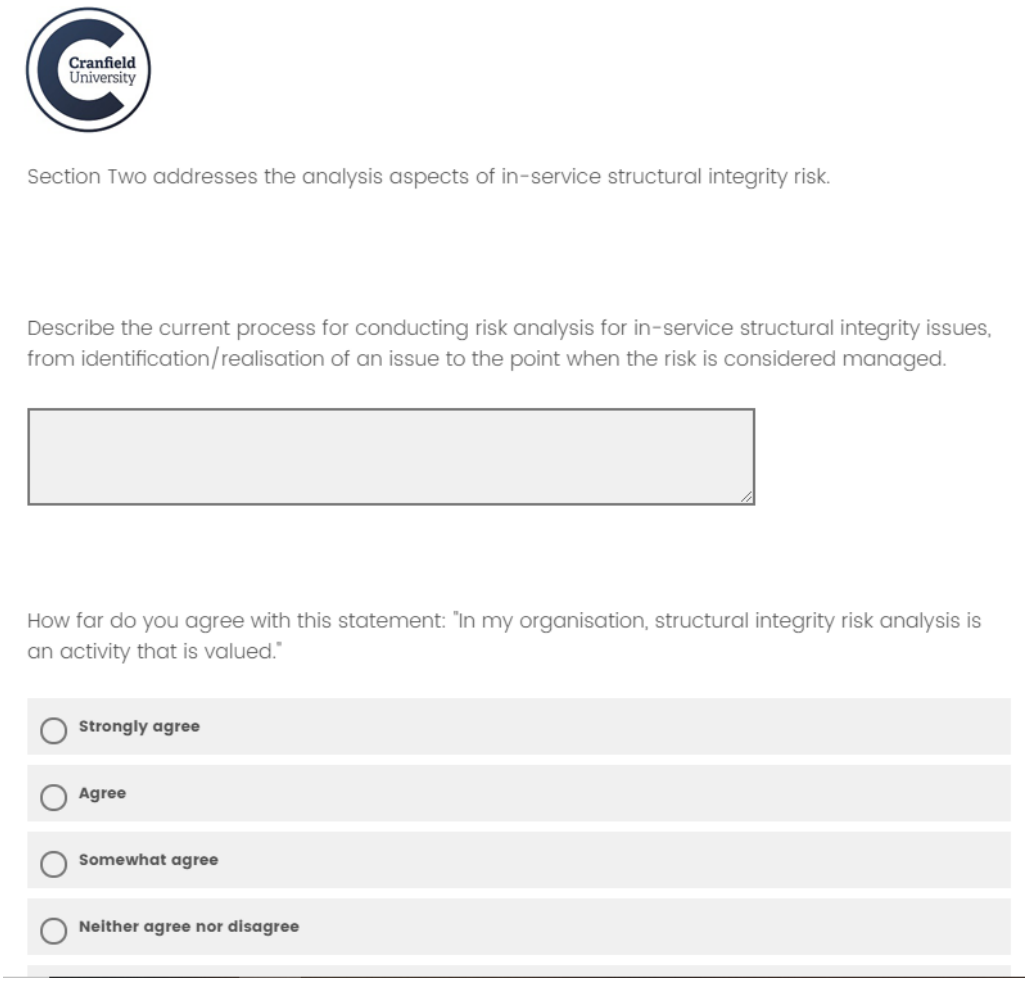


Figure E.3: Practitioner survey screen capture no. 3.

Appendix F

Submitted journal article

Aircraft maintenance decision-making with unsupported judgement under operational constraints

Richard Green[a], Ken McNaught[a,*], Alistair Saddington[a]

[a] *Cranfield University, Defence Academy of the United Kingdom, Swindon, SN6 8LA, UK*

Abstract

In operational situations such as military aircraft maintenance, which experience limitations on time, resource or the information available, it is often not possible to conduct rigorous analysis on complex decision problems. Decision-makers who are compelled to act in such circumstances, may be informed by some level of analysis if available, or else may have to rely on their unsupported judgement. This paper presents three engineering risk decision-making case studies across a 20 year span from the rail, aerospace, and military aviation contexts, highlighting the fallibilities of using unsupported judgements in an unstructured manner. The paper concludes that there is a pressing need to provide some form of support to engineering decision-makers facing operational decisions under severe time pressure. While the ultimate aim must be to improve the quality of decision-making, improved transparency is an important additional benefit. Increased emphasis on decision justification and self-awareness are suggested as potential ways of improving this situation.

Keywords: Decision-making, risk assessment, operational risk, dual process theory, judgement, aircraft maintenance.

1. Introduction

Using technical equipment in operational environments demands quick critical thinking to overcome unforeseen problems given typically limited information, time and resources with which to make decisions. Risk frameworks provide a means to aid decision-makers in environments that involve complex relationships between physical and cyber systems, environmental forces, and social influences. Military aircraft maintenance is an example of such an environment further characterized by many low probability, high consequence events ([1] and [2]). Such environments are challenging for the application of Risk Analysis (RA) at the best of times, particularly given the difficulties associated with the estimation of low probabilities. However, these difficulties are exacerbated when the time available for a decision is short due to operational

*Corresponding author

constraints.

The practical usefulness of RA can be inferred from the proliferation of risk concepts across engineering domains such as oil and gas ([3], [4]), rail ([5], [6]) and aviation ([7],[8]). Increased emphasis on RA is also a response to organisational and human failings found in a number of post-accident investigations (see for example the Challenger Accident [9] and The RAF Nimrod Enquiry [10]). Cross-industry efforts to standardise a lexicon and taxonomy [11] indicate that risk management in technical industries remains a valued activity and policies for best practice are promulgated by domain-specific regulation (e.g. [12]). Probabilistic, quantified risk representations in particular are common in organisational settings and useful for comparing risky prospects and allocating resources [13], encouraging dialogue [14] and justifying the rational basis for a decision [15, p185]. But, critics have claimed that they hide assumptions, require subjective representations of uncertainty and can entail complex computation to generate an output [15, p185].

For military aircraft maintenance decision-makers facing new or unfamiliar structural faults, prescriptive risk assessment requirements are difficult to comply with. Generating a risk measure for a given structural integrity fault is not always practicable in operational situations without relying on judgement. Aircraft structural faults are complex, each aircraft has an individual usage profile, and information is rarely sufficient to model a structural risk meaning that analytical solutions are only available beyond the required time horizon for a decision. Such environmental conditions compel military engineering managers to make decisions by less precise and more subjective qualitative means.

In this paper, three engineering-based case studies are used to highlight the need for improved risk decision support in real-world time-critical operational situations that compel the use of judgement. The constraints that operational settings can impose upon the decision-maker mean that the resulting discussion of the case studies incorporates threads of work from the fields of RA, decision science, rationality, risk perception and applied psychology. We specifically make use of works by Aven [16] and Evans [17]. Furthermore, we support the recent call for more empirically-grounded research in the safety domain [18], and note the prediction that the safe operation of future systems characterized by high complexity and diverse risk sources will require flexible, dynamic interventions with a capacity to handle unforeseen events [19]. The paper first describes the context of military aircraft maintenance risk decision-making (with specific regard to structural integrity), time-critical decision-making and a brief overview of the dual process theory of human cognition. The case studies are reviewed in Section 3, which are real-world engineering-based decision-making situations that provide sufficient evidence to dissect operationally constrained decisions on risky prospects. Section 4 provides the discussion on the observations from the case studies, before the concluding comments are made in Section 5.

2. Risk, decision-making and operational aircraft maintenance

Military aviation operations frequently take place at short notice, in austere locations, with restrictions on in-situ personnel numbers and with extended lines of communication to maintenance support networks. Under these (and sometimes less restrictive) conditions, military aircraft maintenance managers are routinely expected to make 'execution decisions'[1] that balance operational demands for aircraft against the airworthiness risk presented by structural defects. A newly identified crack, material corrosion, accidental damage following maintenance or environmental damage such as a bird strike are typical. Is immediate maintenance to rectify the new fault required, or can it be deferred until later?

Vehicle risk is typically an *average* risk level established during design and certification, dependent upon assumptions and uncertainties about real-world usage and maintained during service in accordance with regulatory requirements[2]. New and unexpected faults can create doubt regarding a priori design beliefs, undermining existing risk assessments. Faults must be managed appropriately to balance the risk of harm and equipment availability requirements, since operational success may be dependent upon the immediate risk assessment and subsequent decision. This practical observation is consistent with leading thinking in safety science, which identifies that complexity across sociotechnical boundaries that influence safety decisions make complete foresight of safety conflicts impossible, necessitating a more resilient approach [19].

Understanding risk helps decision-makers construct rationales, form logical decisions regarding intervention/ mitigation activity and provide reassurance to the stakeholders of aircraft safety. In domains where accountability and decision transparency are important (required by UK legislation [22] and indicated in military regulation [23]), it is appropriate to understand how a decision-maker has selected their course of action. However, if a decision-maker's assessment of risk is "normally invisible during the decision-making process and an informal assessment process is concealed in the mental models and the experience of professionals"[20, p117], the route to the outcome becomes opaque. The cognitive aspects of decision-makers should be accounted for in designing a risk management strategy.

Time and effort required for a requisite decision analysis grows with the

[1] *Execution decisions* are "made by sharp-end personnel with minimal or no planning, typically during the implementation/execution of different work activities" [20], for example to follow or deviate from procedures, or ignore or react to deviations from normal working conditions.

[2] European Union Aviation Safety Agency regulations require operating risk to be expressed as a quantified average of less than one catastrophic event in 10 million flying hours [21, p46] and no worse than one catastrophic event in every 2 million flying hours [21, p47].

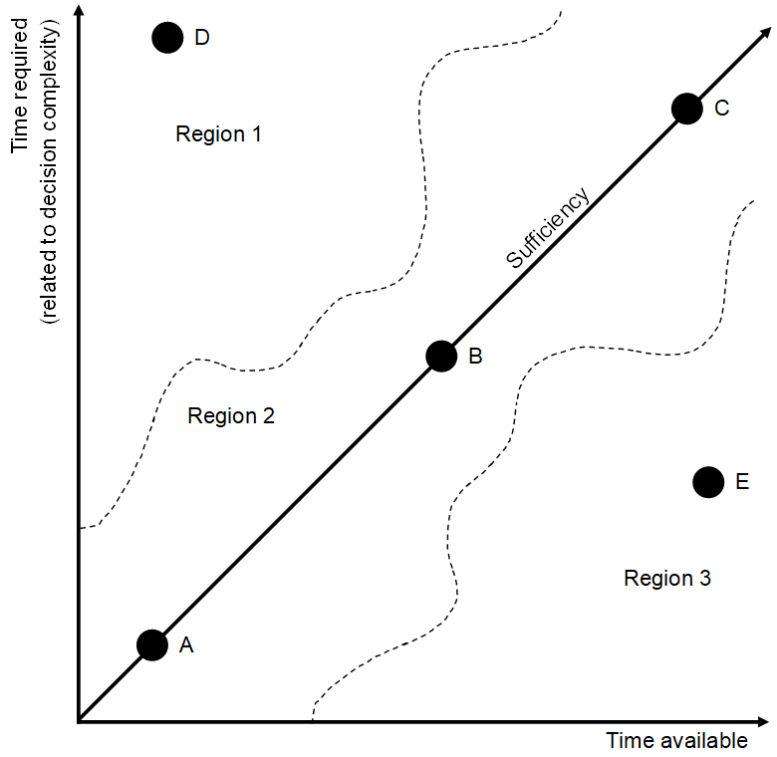


Figure 1: An extended decision-making hierarchy, adapted from Howard and Abbas [24]. The line of sufficiency in Region 2 corresponds to decision situations where the time required to make a requisite decision is equal to the time available.

complexity of the situation, shown as a decision hierarchy in Howard and Abbas [24, p28]. Aircraft maintenance situations can restrict the freedom to conduct an appropriate analysis for a complex decision - decisions worthy of rigorous analysis must, instead, often be made with very limited resources and within a tightly constrained decision horizon[3]. Figure 1 categorises situations with varying levels of time required to reach a requisite decision (which can also be associated with decision complexity) and varying levels of time available to reach the decision. It extends the decision hierarchy of Howard and Abbas [24]

[3]The decision horizon is the time within which the decision is necessary, influenced by any environmental, societal, political, operational, or other demands on the decision-making system.

to indicate regions of varying time sufficiency above and below the threshold of sufficiency represented by the straight line in region 2. Points A, B and C represent the relative positions of the three layers in the existing decision hierarchy and correspond to different levels of decision complexity. In region 1, the time required for a requisite decision is greater than the time available – typically decisions with high complexity that, nonetheless, must be arrived at quickly. Point D might represent an urgent operational decision regarding whether an aircraft tasked to a mission with strategic importance should fly its next scheduled sortie when damage to critical structure has been identified during preparation of the aircraft. In region 3, the time available to reach a decision is greater than the time required to reach a requisite decision at that level of complexity. While an efficient decision-maker will not actually use this excess time, an inefficient decision-maker might. This could manifest itself as a decision situation being over-complicated - typically, low complexity decisions being afforded more time than is necessary. Point E might be a minor decision such as how often to conduct management assurance checks on engineering processes and activities. How much of the extra time available is actually used is related to the decision-maker's efficiency. Extending Howard and Abbas [24] in the form of Figure 1 is helpful *a priori* as a guide to stakeholders to provide consistency of expectation, and *post hoc* to reflect on a past decision and identify whether constraints could have been relaxed, for example.

With regards to aircraft structural damage specifically, the most common forms of damage identified during aircraft maintenance are fatigue and corrosion, as indicated in Table 1, although these mechanisms can also be triggered by human events such as maintenance mishaps (e.g. accidental damage from tooling) or environmental events such as climate exposure (e.g. from deploying aircraft to maritime or coastal environments). Structural degradation is further complicated in that two aircraft of identical design and intended usage may vary individualistically as a result of natural variation in material or unique events such as high loading events (e.g. turbulence, heavy landings). Damage observed on one aircraft cannot be automatically assumed to be common across a fleet. Given sufficient time and resource, structural damage can be modelled. Methods such as finite element analysis can take into account the effects of expected loads exerted on the aircraft structure for future flights. These can provide probabilistic interpretations that can be used in risk calculations.

Quantitative methods for assessing structural integrity risk are largely simulation based, dynamic in nature (reflecting the temporal aspects of structural integrity), built on statistical theory, and require detailed knowledge of material degradation relationships and key structural system variables (see for example [26–34]). However, such methods are information hungry, require expertise to build the model and appropriate software to conduct simulations. These conditional requirements are unlikely to be fulfilled under operational conditions that limit information, time, or resources available to a decision-maker compelled to make a time-critical choice on aircraft airworthiness with the

	Percentage of failures	
	Engineering components	Aircraft components
Corrosion	29	16
Fatigue	25	55
Brittle fracture	16	-
Overload	11	14
High temperature corrosion	7	2
Stress corrosion cracking/corrosion fatigue/HE	6	7
Creep	3	-
Wear/abrasion/erosion	3	6

Table 1: Frequency of structural failure mechanisms in engineering and aircraft components, reproduced from [25].

people, knowledge and equipment available at their immediate disposal. Even if the situation allows quantitative methods to be used, the subjectivity applied in building a risk model [16, 239] can lead to problems in validation and accuracy in the quantitative risk assessment [14]. Combined with the individualistic nature of material degradation on aircraft, analytical structural risk methods entail a significant information burden and are unlikely to be suitable for use in time and resource-critical operational situations.

A further consideration is that engineering maintenance managers are typically generalists. They are highly skilled in balancing the time, resources and information available to conduct maintenance across a range of aircraft systems (such as avionics, mechanical systems, weapons systems, structural items and electrical components) and multiple aircraft, with the operational requirements of the equipment. However, lacking the skills and knowledge required to conduct structural analysis of material failure, maintenance management reliance on experts is expected and necessary. But even experts are dependent on judgement and will suffer human biases or heuristic influences when assessing risks, even in analytical assessments ([35–37]). In an operational situation where the decision basis will not have the support of rigorous analysis, UK military aviation regulation permits the use of judgement:

"On occasions when a Maintenance activity cannot comply with relevant TI[4], or there is insufficient resource, the Maintenance must remain incomplete. However, an operational requirement may necessitate a Maintenance activity being completed prior to resources becoming available or prior to an approved

[4]Technical Information, such as maintenance procedures or special instructions.

and promulgated TI amendment being issued by the TAA[5]*.....Deferring Maintenance and deviating from TI carries risk. When considering deferment or deviation, the authorized individual must assess the associated risks and consider all factors that will mitigate the risk and ensure the Air System is airworthy. The mitigating factors must be adequately documented in the appropriate Maintenance work order."*[38, p10].

Consideration of risk when applying such policies may be quantitative or qualitative, which the above extract leaves to the preference of the decision-maker. Alternatively, taking an accident avoidance approach away from precise risk statements may be preferred [20, p119]. However the decision is formed, the judgement of a decision-maker may be misguided by cognitive influences: various biases leading to a wrong conclusion; social factors and operational pressures may distort the urgency; individual risk appetite may promote overly conservative action resulting in a suboptimal use of resources. Ultimately, errors of judgement could expose an aircraft and its occupants to unintentionally high levels of risk. Risk assessment should help prevent such problems by logically structuring available information about uncertain future outcomes to enable a rational decision basis. But under operational constraints, judgements about risk that lack analytical support could be challenged on the grounds of credibility, irrespective of the decision alternative selected. For instance, a maintenance manager's decision to ground an aircraft based on unsupported judgement may invite dissent and criticism from operators, availability managers, or other stakeholders; the manager would be expected to justify their argument.

2.1. Dual Process Theory

Dual process theory is a psychological theory of how humans think and is useful for considering unsupported judgements about risk. The theory claims that there are two cognitive means by which an individual processes information: a fast, intuitive and experiential reaction (known as Type 1 or System 1); and a slower, analytical, logical and deliberative mode (known as Type 2 or System 2)[6]. The characteristics typically associated with each are summarised in Table 2. Evans [17] opposes a drift in the functional application of Type 2, from supporting and rationalising Type 1 thinking, to engaging reasoning to draw conclusions and replace Type 1 thinking. He argues that the latter

[5] Type Airworthiness Authority, a UK Ministry of Defence organisation with responsibility for the overall airworthiness and resourcing of a fleet of aircraft. They often act as the intermediary between the operators of the aircraft (the UK Ministry of Defence) and the manufacturers (defence industry), but provide immediate fleet and individual aircraft airworthiness advice to operators.

[6] Some authors, such as Aven [16] apply the 'System 1/2' lexicon, Evans [17] discusses why this vocabulary, which developed from an evolution of early Dual Process Theory discussions, is not helpful because Type 2 processes can be formed from multiple thought mechanisms or systems such as reasoning, planning, imagination, and strategic thinking. All have a specific function but remain a Type 2 response.

Type 1	Type 2
Experiential	Analytic
System 1	System 2
Subjective	Objective
Hot emotional	Cold rational
Intuition	Analytical
Automatic	Deliberative
Non-verbal	Verbal
Animalistic	
Narrative	
Fast	
Natural	

Table 2: Variations in the lexicon used in literature to describe Descriptive and Normative Decisions ([39], [16]).

abstraction supports a misconception that Type 1 is generally biased and Type 2 is normatively correct, and highlights differences in opinion as to how Type 1 and Type 2 processes interact - in series or in parallel ([17, p387], also see [39] and [40, p113]).

One unifying proposal is The Default- Interventionist model [17] (shown in Figure 2 as the solid lines). This model applies a widely-held belief that Type 1 processes operate much faster than Type 2 processes; the latter engages the working memory and can overwrite Type 1 thinking if appropriate [41, p10]. The two modes are not mutually exclusive and the degree of interaction between them is influenced by factors that are both individualistic and environmental. The dashed lines illustrate how the Default-Interventionist model may be extended to represent unsupported and supported judgements, discussed further in section 2.2.

Dual Process Theory and Risk. Although making a welcomed bridge between dual process theory and risk, the distinction between Type 1 and Type 2 thinking made in Aven [16] is inconsistent with that defined in Evans [17] and Over [41]. For instance, it is suggested in Aven [16] in his example of a bus bombing that poor knowledge about the frequency of a historical event triggers a Type 1 response, enabling avoidance of the risk by not getting on a bus. Similarly, his approach to incorporating Type 1 thinking into professional RA focusses "on the basic features of risk and the issues on which System 2 thinking is conditional, i.e. K" [16, p241], where K is the knowledge state of an individual. Reflectively processing historical event data and acting upon conditional information are

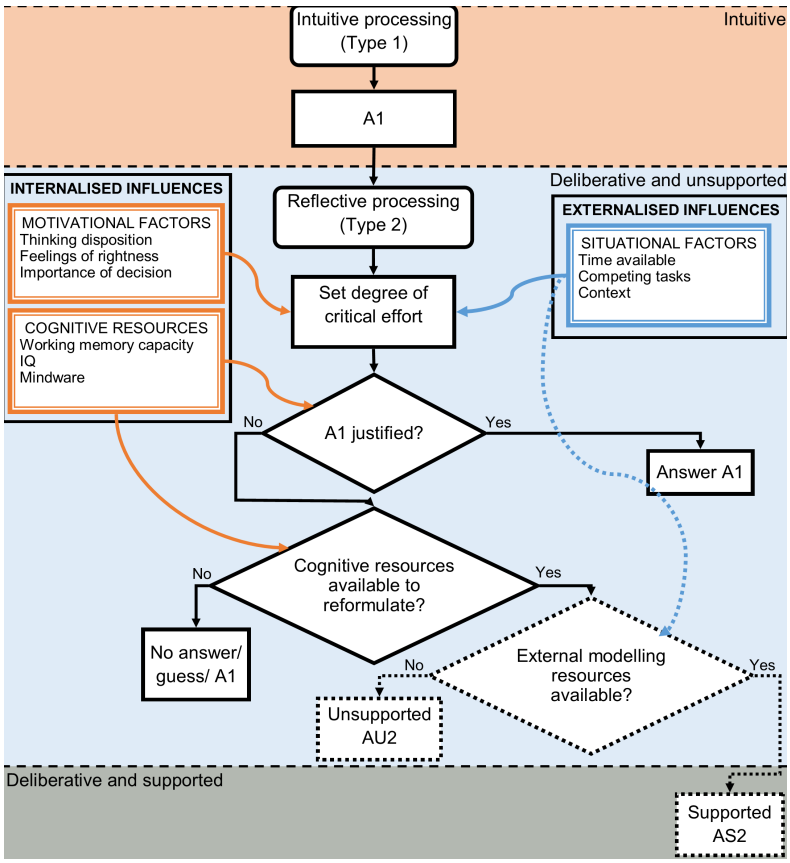


Figure 2: The Default-Interventionist model (solid lines) including extension by the authors (dashed lines), describing how individuals form decisions from the position of a Dual Process Theory perspective [17, p395].

deliberative actions - Type 2 responses. While this correction may appear semantic, accuracy in definition ensures consistency in understanding across the readership and for further development work.

According to Aven [16, p242], traditional risk methods are too narrow in their consideration and do not account for holistic perspectives that are formed from individual experiences. These may constitute the intuitive processing associated with Type 1 thinking. Certainly, substantial literature on cognitive

biases shows how human cognition can be misled (for example [42] and [43]), effects that would be expected to be minimised through the RA process. Of course, and as many cognitive bias experiments have shown, humans also have the potential for making errors when using Type 2 thinking, e.g. to estimate probabilities, with obvious repercussions for risk analyses: errors of competence; errors of comprehension in applying norms or paradigms; and deep rooted misconceptions when making judgements on probability [44, p100-104]. Scope for cognitive error is also plainly apparent through the use of analyst and expert subjective judgements in traditional risk methods [16, p242-243] through their choice of parameters or modelling criteria, for example. Holistic subtleties such as reversibility, persistence, ubiquity, tolerability, equity, catastrophic potential, controllability, or voluntariness [45, p442] and cognitive characteristics like "affect"[7] therefore have bearing on the conduct of risk assessment, decision-making, communication and management: "...while we may be able to "do the right thing" without analysis (e.g. dodge a falling object), it is unlikely that we can employ analytic thinking rationally without guidance from affect somewhere along the line. Affect is essential to rational action" [39, p314]. Aven [16] mechanised this thinking in a checklist approach using discrete steps to extract the background knowledge and make qualitative judgements on key risk issues, to "lift the status and authority" of Type 1 processes [16, p242]. Although, as we discuss above, his method includes deliberative Type 2 thinking.

Making a distinction between whether a particular thought is Type 1 or Type 2 in origin is not necessarily important, since for decision-makers "their concerns are the "unconditional risk" (C, U)[8] and not the conditional analyst risk characterisation $(C', P|K)$[9]" [16, p242]. Improving the justification of judgement and beliefs benefits the epistemological footing of an argument [16, p242] and supports decision-makers who are deliberating risky prospects without the luxury of analytical support.

2.2. Shaping the risk assessment around the decision-maker

Extending the Default-Interventionist model to the aircraft maintenance risk decision context requires distinctions to be added to account for the practical resource and information limitations enforced upon the decision-maker. These additions are shown in Figure 2 by the dashed lines. For aircraft structural integrity specifically, these additions account for whether the decision-maker can be supported by trustworthy structural analysis/expert information or not. Situations that have the resources, information and time to enable the formation of analytical assessments are referred to here as a "supported" Type

[7] Affect means "the specific quality of "goodness" or "badness" (1) experienced as a feeling state (with or without consciousness) and (2) demarcating a positive or negative quality of a stimulus" [39, p312].

[8] C are the actual consequences, U is the actual uncertainty.

[9] C' are the consequences addressed in a risk assessment, P is a measure of uncertainty, K is the background knowledge that supports the judgements of C' and P.

2 risk assessment. Under operational constraints, however, an “unsupported” condition may arise when decision-makers are compelled to form a decision but their judgement lacks the backing of an analytical assessment. A deliberative assessment may still be possible, reliant upon the decision-maker’s ability to act on the available information. The labels ’AS2’ and ’AU2’ in Figure 2 refer to the answers produced after following a supported or unsupported Type 2 reasoning process, respectively. These situations are also depicted in Figure 3 and Figure 4, which illustrate the (unscaled) time of Type 1 and 2 triggers occurring, and the time periods these modes of thinking operate over. Figure 3 depicts situations that arise when a Type 2 risk assessment has been able to make use of modelling, analysis, deterministic solutions, historical information, or another credible and rational tool appropriate to a given decision. Figure 4 shows the unsupported condition, which lacks the use of such reference materials (labels are consistent with Figure 3).

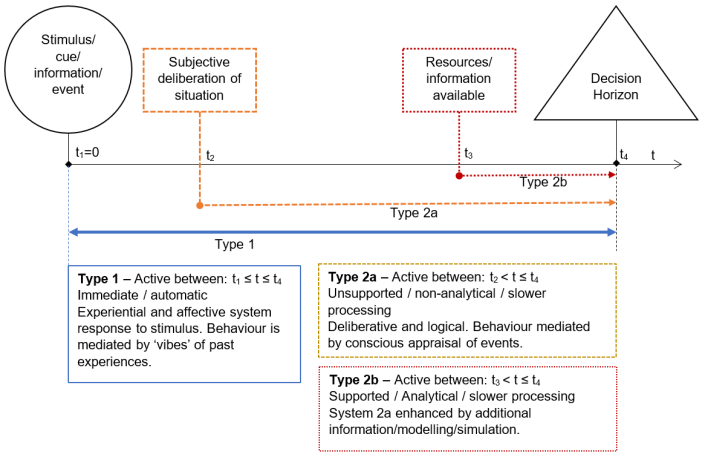


Figure 3: Decision follows a ‘supported’ Type 2 thought process.

Having no analytic support affects the extent to which an individual may rely upon their intuition or ‘gut feeling’ about a particular maintenance problem and influences the quality of the Type 2 thinking process. While Aven [16] calls for Type 1 thinking to be included in risk assessments, it is already there. According to the dual process model, it is already influencing Type 2 deliberations, whether supported or not, which form the basis of formal risk assessments. Not withstanding the management strategies that already exist to cope with such situations (e.g. the cautionary principle [46, p29]), in certain situations unsupported Type 2 decisions become a necessity.

Refocussing risk assessments to be cognizant of the situationally-specific

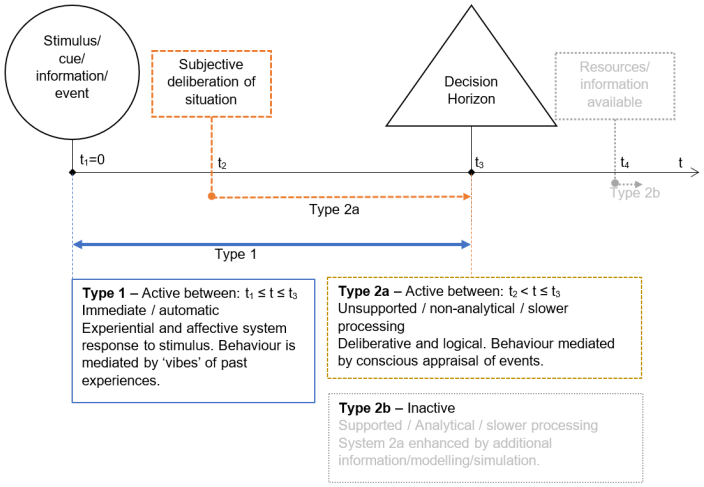


Figure 4: Decision follows an 'unsupported' Type 2 thought process.

capabilities of the decision-maker moves away from traditional engineering objective approaches that focus on hazardous events, reliability and material performance expectations. Human reactions to an emerging risk create automatic and uncontrollable perceptions of whether to consider the situation threatening: "what can happen, the potential consequences, judgements of likelihood and the knowledge base on which these judgements are based" [16, p239]. Such judgements are deliberative, Type 2 in nature but unsupported. Systematically accounting for these holistic subtleties of analysis outwith likelihoods and quantifiable effects may enable proactive sensing of failure signals or encourage risk questioning [16, p243]. This could benefit risk activities by transcending disciplinary boundaries and provide more consistent and coherent scenarios for decision-makers [45, 442].

A consensus definition of risk already identifies that consequences impact something that humans *value* [11]. This move has been a topic of discussion in several subdomains. Hansson and Aven [47] suggest that the information flow in a decision-making scenario for risk includes the values of the decision-maker. van Asselt and Renn [45, p442] champion an holistic approach to "framing, appraising, characterizing, evaluating, and managing risk" and argue for more coherence between risk assessment and risk management in their discussion on risk governance. Pidgeon [48] debates the need for incorporating both scientific and value statements to describe the 'tolerability' of risk.

With the unsupported and supported decision constructs in mind, this paper seeks to highlight how unsupported Type 2 thinking influences engineering risk decisions in situations that restrict the support tools available to the decision-maker. We believe that the use of unsupported decisions and experiential-based thinking is more common place in real-world decision-making than might be otherwise assumed and that it deserves more attention in the risk literature.

3. Is unsupported thinking used in operational engineering judgements of risk?

3.1. Method

By analysing narratives of engineering incidents, evidence of unsupported risk judgements can be extracted or inferred from the activity. The cases presented show operational situations that required engineering management decisions on risky prospects. Although our interest in this area is primarily motivated by military aircraft structural integrity, case selection was not limited to this domain. This helps to demonstrate generality but also broadened the available sources from which to select suitable candidates to study.

Case study selection. Candidate cases were selected as meeting the following situational criteria:

- A risk emerging within an established risk management system.
- Operational environmental constraints.
- A human component that selects whether intervention is necessary.

The naturalistic evidence studied here builds upon the generic examples seen in [16]. Structural and engineering-focussed cases that involve human perceptions of risk within complex engineered systems were favoured. The narratives used provide sufficient grounds to infer that decisions were based on Type 1 or unsupported Type 2 thinking within an established organisational risk management system, where a need for transparency is necessary. Such requirements are typical in public organisations, where expenditure is subject to public scrutiny. The cases necessarily describe decision-makers in situations with real-world limitations lacking the time, resource, information or cognition to conduct a full analytical risk assessment. The cases also span a 25 year period to illustrate the current norm and historical pedigree in risk assessment practice.

Numerous cases were found that illustrate where hazards have emerged and created risk to operating equipment. Harder to identify were documented cases that detail where and how the human system has interacted with the risk management system through the use of judgement. There were also many cases that describe the technical situation and available information, but few that explore the intervention choices and justification made by the human component. Discussion regarding the suitability of the sources used is provided with each case study.

Code	Description	Reference
[LA]	Learned associations between ideas	[16, p238]
[PERC]	Perception of risk as high, fearful and unacceptable	[16, p238]
[RHI]	Judgements of the risk as high	[16, p238]
[RTHI]	Judgements of the risk as too high	[16, p238]
[UNHI]	Judgements of the uncertainties as high	[16, p238]
[UNTHI]	Judgements of the uncertainties as too high	[16, p238]
[NBPA]	When judgements of how large the risk is are not based on a professional assessment of how large the risk based on either: historical data, probabilistic analysis, or by comparing with other activities	[16, p238]
[BGKN]	Use of unsupported thinking in professional judgements, such as risk aspects that may be hidden in K (background knowledge) like assumptions.	[16, p239]

Table 3: Expressions used to codify the case studies in identifying unsupported decision judgements to risky situations.

Case study review methodology. A qualitative review based on textual accounts of historical events has been conducted using a template analysis approach [49]. This approach extracts a thematic structure from [16], shown in Table 3. The codewords defined do not differentiate between Type 1 and Type 2 thinking because this does not contribute to the focus of the study. The main intent is to illustrate the use of unsupported risk decision-making by using [16] as a descriptive framework for testing the selected cases against. [NBPA] modifies the source text to highlight the distinction between unsupported judgements and those based on professional analysis[10]. Where these codewords seem applicable, the relevant section of text is annotated with the codeword. If a code is applied to a section of text and there is no evidence or inference to a supported assessment, this is considered as unsupported thinking.

[10]Original source reads “Professional judgements of how large the risk is, based on historical data and probabilistic analysis” and “Professional judgements of how large the risk is, by comparing with other activities” [16, p238].

3.2. Case 1 - Lamington Railway Viaduct Structural Failure

Lamington railway viaduct partially collapsed following erosion of the river bed beneath the structural supports of the viaduct. Related risk information is traced back nearly 30 years to when another railway structure in the UK collapsed due to erosion of the riverbed beneath supporting structure, highlighting organisational failings to appropriately handle a known risk. The Lamington viaduct incident report was published on 14 Nov 2016, taking approximately 11 months to investigate, conclude and release the findings.

3.2.1. Limitations of the sources used

The incident report seeks to improve railway safety without establishing blame or liability [50, p3]. Accordingly, the textual summaries and inferences of the actions can be considered factual as far as was investigated at the time. However, as a single source from one organisation's perspective, investigators may have not had access to particular evidence, the report may suffer individual bias (on the part of the investigators), or they may have been subject to legal constraints at the time.

3.2.2. Narrative of events

On 31 December 2015, prolonged heavy rainfall and subsequent high flow in the River Clyde, Scotland, UK, created the conditions necessary for the erosion and removal of river bed material under the base of the Lamington viaduct. The erosion phenomenon known as 'scour' caused subsidence of the viaduct support structure, which became noticeable by train drivers passing over the viaduct that morning as a 'dip' in the track. Attending track maintenance engineers initially diagnosed minor track deformation as the cause for the reporting 'dip', but closed the line when a passing train created surprisingly large track movements. By this point, several trains had been permitted to travel over the viaduct. Subsequently, the maintenance team found damage to the central pier of the viaduct, which was the root cause of the track deformation. The trains that had passed over the damaged viaduct had been fortunate to escape derailment or to cause a more dramatic collapse of the viaduct. The incident was declared a "dangerous occurrence" [50, p26].

The viaduct had been subject to multiple formal assessments since 2005, consistently scoring as a 'high risk' structure by a contracted surveyor using a network-wide scour priority rating system. Annual underwater inspections provided routine mitigation activity and extreme weather precautionary measures that provided reactive procedural mitigation (such as monitoring river levels and closing the line if a threshold waterline was reached), but these had fallen out of use owing to organisational changes. Planned scour reinforcement works to the viaduct were deferred from 2015 until mid-2016 owing to essential environmental approvals not being secured.

3.2.3. Unsupported judgement decision points

Two decision points highlight the use of unsupported thinking within Lamington's scour risk context.

The on-call Track Maintenance Engineers close railway. On arrival at Lamington to assess the reported track fault, the maintenance team's combined experience initially misdiagnosed minor 'dips' in the line. This deliberative conclusion led to the exposure of three trains to a structurally-defective viaduct. The decision may have been reached because the maintenance team believed (or assumed) the track dips were isolated and the most likely root cause from previous experience or historical evidence: "the vast majority of track faults are directly related to track condition" [50, p35] ([LA], [NBPA] or [BGKN]). The maintenance team's lack of knowledge regarding scour risk and the specific threat posed to Lamington [50, p20] was also a missing cue. The decision was unsupported, made on the basis of underlying assumptions being true [BGKN] such as the fault being track related only, the supporting structure being unquestionably sound and with no other external influencing factors. The grounds for these assumptions were then found to be false once the maintenance team identified the structural damage to the viaduct and reacted by closing the line immediately. The team were not bridge specialists (having requested the attendance of a bridge examiner and being unaware of scour risk) but perceived the risk to be high, fearful or unacceptable [PERC].[11]

The non-structural-specialist track maintenance engineers were required to make a timely decision on an unfamiliar risky prospect, which was actually known about within the organisation's risk management framework. Their decision was unsupported by existing formal RA and required the use of their judgement to escalate the situation for wider organisational consideration.

Scour reinforcement works. Two contracted formal initial scour risk assessments (in 2005 and 2010) classified Lamington as at high risk of scour. Recommendations to conduct a follow-up detailed assessment and regular physical scour and flood damage checks were only partially complied with. Mitigating annual underwater inspections only provided 'point-in-time' conditional monitoring of the viaduct support structure, and a necessary repair took 3 years to install and was temporary in nature. Despite knowing the formally assessed risk, management elected not to apply a full and timely risk intervention [NBPA]. Lamington's listing on a schedule for scour protection works in 2012 indicates

[11] It is noteworthy that railway signalling and control room staff would not have been permitted to allow trains to be exposed to *suspected* structurally damaged equipment [50, p35], however control room staff were also unaware that Lamington viaduct had been assessed as being at high risk of scour and therefore susceptible to structural damage. Although this suggests an information and communication breakdown, it also demonstrates the conservative intervention required in the face of insufficient structural risk information ([51])

that the risk influencing factors were known about:

"Lamington was prioritised due to its location on the West Coast Main Line and known shallow foundations. [50, p33]"

This shows business and reputational concern for the importance of the route that the Lamington viaduct was situated on ("its location on the West Coast Main Line") and knowledge of the formal risk assessment findings ("known shallow foundations"). A detailed scour risk assessment was conducted in 2013, again scoring Lamington as high risk. The decision-makers have used the RA as a constituent in a wider value-focussed decision-making process ([47]), since rail authority personnel "requested authority to defer the scheme until the next financial year due to environmental permissions not being sought in time for the work to be completed during the low flow summer period" [50, p33] [NBPA].

Decision-makers did not perceive the risk (or uncertainty surrounding the formal assessment conclusions) from scour to be sufficiently high to cross their personal threshold for intervention ([RHI] and [UNHI]). Decision-makers may have been mistakenly optimistic from the annual underwater inspections and overly reliant upon assumptions regarding the adequacy of risk mitigation that was in-place. When scour reinforcement works were deferred in 2015, the stated reason was that there were "no new defects noted by the latest visual examination" [50, p33]. This would be reliant on assumptions such as observable gradual deterioration preceding a total collapse but not deteriorating catastrophically between inspections [BGKN]. Other influences on this judgement may have been perception of the true risk to Lamington, assumptions hidden in some personal knowledge of the structure in place at Lamington, the behaviour of the River Clyde, or of the mechanics of scour risk ([NBPA] and [BGKN]). Decision-maker judgements were made *in spite* of the formal analytical risk assessment and there is no evidence of the grounds for why robust intervention activity was so tardy.

The inability to substantiate structural risks and prioritise mitigation activity highlights the difficulty that decision-makers have in communicating the urgency of risk intervention requirements on structure to resource managers and budget approvers. Meaningful intervention was only initiated once a clear unsafe threshold was crossed - when damage was visible. Despite quantitative assessment scoring Lamington as a high priority, qualitative value-based judgements rationalised continuing deferment of intervention activity. This demonstrates unstructured use of unsupported judgement in spite of an available analytical risk assessment.

3.3. Case 2 - The Challenger Space Shuttle Accident

The second case concerns the well documented Space Shuttle Challenger accident in January 1986 (mission number STS (Shuttle Transport System) 51-L), which experienced a mid-launch complete vehicle destruction. The

mechanical cause was found to be the failure of O-ring seals to contain hot propellant gasses in the solid rocket booster (SRB), a major component that was designed to be recoverable following a launch, refurbished and reused in subsequent launches. The accident was preventable, and the post-incident investigations found that a series of conflicts between engineering and management decision-making led to a collective decision to launch, despite initial recommendations of 'no launch' and sustained dissent from specialist engineering personnel. In the wake of the accident, a Presidential Commission was convened [9] to investigate the accident and all staff involved were subject to scrutiny by both the commission (under oath) and the public owing to media coverage [52, p388].

The case illustrates a major technical project that operated a risk management system and was subjected to operational stress factors for the Challenger launch scheduled for 28 January 1986. Although there is substantial discussion on this case that identifies organisational safety failures back to the design choice of the Shuttle vehicle, unsupported thinking is widespread in the hours prior to the launch of STS 51-L. For brevity, only one circumstance is explored to demonstrate where the shuttle programme risk management system might have included unsupported Type 2 thinking in the assessment of true risk.

3.3.1. Limitations of the sources used

The secondary sources used are contrasting narrative accounts of the decision to launch the Challenger shuttle. The Presidential Commission Rogers Report is criticised for being too focused, misunderstanding NASA procedures in decision-making, and not including some personnel's testimony in their consideration [52, p59], suggesting it is narrow in its conclusions. But it provides timely and personal testimony from the protagonists involved in the flight safety decision process on 27 and 28 January 1986, which is the prime consideration in studying how far unsupported thinking contributed to the Challenger accident. To balance the use of the commission report, [52] provides an alternate perspective and source of wider testimony not included in [9].

3.3.2. Narrative of events

Following Challenger's planned launch on 27 January 1986 being deferred owing to strong winds at the launch site, the launch was re-planned for 09:38 on 28 January. The forecast temperatures were colder than any previous launch. SRB manufacturing engineers at Morton Thiokol (MTI) and Marshall (responsible for maintaining oversight of the SRB contract among others) raised their initial concerns of the temperature effects on the O-ring seals between the joints of the rocket segments shortly after the decision to defer the mission 24 hours

([9, p87] and [52, p286])[12]. These initial concerns triggered activity by both MTI and Marshall to collate the available data regarding temperature effects on SRB joint sealing, which was then discussed during two teleconferences between MTI, Marshall and NASA the night before the launch. During the second and more widely attended of these teleconferences, MTI initially presented a formal 'no launch' recommendation. Marshall rebutted the rationale on the grounds of inconsistencies between the data they presented and a history of behaviour by MTI who had not rejected launches given similar forecast temperatures[13] to those predicted on 28 January 1986 [52, p155, p308-310]. Consequently, Thiokol reviewed their rationale and returned to the conference half an hour later with the conclusion that despite concerns over the cold effects on the O-ring seals, the data was inconclusive. MTI subsequently signed their telefax recommendation for the launch of STS 51-L on 28 January 1986. Challenger launched at 11:38 on 28 January 1986 at an ambient launch pad temperature of 36°F [52, p7], exploding 73 seconds later in a fireball and with total break-up of the vehicle. The seals were found to have failed to perform their role during the first seconds of launch.

3.3.3. The initial 'no launch' rationale

The initial concerns raised about the cold by MTI and Marshall personnel infer an holistic, deliberative consideration regarding the primary cue (the forecast ambient temperature at the launch site) and the existing O-ring sealing risk. The O-ring seals had been subject to launch constraints and individual component testing after erosion and blow-by of the O-ring seals had been spotted following previous missions[14]. However, a normalisation of deviance within the team meant they had learned to accept some erosion of the O-ring seal during launch. The fact that forecast temperatures prompted individuals to behave differently and with urgency for STS 51-L indicates that background assumptions (that were perhaps based upon on the *normal* conditions expected for a launch) had become challenged. The uncertainty regarding how the Shuttle's SRB seals would perform as a system, given the unprecedented cold, influenced multiple personnel to consider that the risk was high given their knowledge of previous instances of erosion and blow-by ([PERC], [UNHI] and [RHI]):

- Marshall's Larry Wear, who was the initial trigger for consideration of the risk from cold temperatures, recalled the effect of low temperature on STS

[12] MTI's design of the SRB required the rocket to be manufactured as several discrete sections at their plant in Utah, USA in order to enable transport by rail to the launch site in Florida, USA. This logistical constraint meant that the SRB could not be manufactured as a solid single (monolithic) section. A monolithic design was simpler by comparison, which had been offered as a solution by competing contractors when the contract was awarded [52, p425]

[13] Mission STS 51-C was scheduled for launch on 23 January 1985, but was delayed by the Mission Management Team owing to the below freezing temperatures. No concerns had been raised by MTI or Marshall regarding the effect of cold temperatures on propellant seal integrity.

[14] STS-2 was the first indicator of erosion, heat effects were found on two rings on STS 6 [52, p149], and then in 1985 seven out of nine shuttles launched experienced erosion and/or blow-by, with the worst arising yet seen when STS 51-C's blow-by reaching the secondary O-ring [52, p153]

51-C in January 1985 and asked Boyd Brinton (MTI), to call Thiokol's Utah plant and find out if there were any concerns [52, p286].

- Team manager for the Solid Rocket Motor (SRM) project, Robert Ebeling (MTI), responded to the request from Brinton by calling a meeting of subject matter experts to review his initial temperature concerns: "The meeting lasted one hour, but the conclusion of that meeting was Engineering...were very adamant about their concerns on this lower temperature, because we were way below our data base and we were way below what we were qualified for." [9, p87]

- Allen McDonald (MTI) was at the Kennedy Space Centre: "I took that data [the forecast temperature data] and called back to the plant and sent it to Bob Ebeling and relayed that to him, and told him he ought to use this temperature data for his predictions, but I thought this was very serious and to make sure that he had the vice president, engineering, involved in this and all of his people; that I wanted them to put together some calculations and a presentation of material." [9, p87]

- Other Marshall employees also formed subjective qualitative-based concerns about the anticipated temperatures on the performance of the O-ring seal:

  "Schell and Riehl concluded that the O-ring would be all right at 25 °F because when the ignition pressure hit it and jammed it into the gap, it would seal. Schell said: "It would have been a little harder, there's no question about that. I mean, there's data all over the world to show that it would have been a little less resilient and a little harder, but at those pressures, it would have sealed." "[52, p290]

In summary, uncertainty regarding the consequences of the weather compounded existing primary and secondary o-ring resilience concerns given the launch dynamics of the SRB joints, the ignition pressure and the time taken for the o-ring to form an adequate gas seal. [52, p290]. There was uncertainty owing to the lack of scientific data regarding cold effects on SRB joint sealing capability in full scale equipment configuration, and a conflict in the conclusions drawn from component test results[15] and historical launch

[15]The team had data on subscale component tests of resiliency (prevention of blow-by) to temperatures down to 30 °F, had conducted sufficient erosion resistance tests to conclude that the seal would still perform beyond the worst case seen from a live launch, and had data on how the O-ring seal material hardened with reducing temperature.

data[16]. Despite Thiokol's worries about the cold, no 'Launch Commit Criteria'[17] had been established for the O-ring temperatures after STS 51-C in January 1985 because "it was nobody's expectation we would ever experience any cold weather to that degree before we had a chance to fix it again" (Roger Boisjoly, MTI) [52, p308]. Consequently, the only full scale vehicle temperature-related data of the SRB performance was perceived to show no correlation between temperature and the O-ring erosion and blow-by (see footnote 16).

MTI were unable to support their 'no launch' argument with any analytical evidence because there wasn't any for the whole SRB system at the forecast temperatures. MTI used unsupported judgement to argue that cold temperatures had resulted in O-ring erosion on previous flights. However, because available o-ring performance data was uncorrelated (footnote 16) Marshall and NASA were able to construct a convincing counter-argument. Notably, the influence of group dynamics on the decision made, is discounted by Vaughan [52, p404]. Regardless of the counter-argument, MTI engineers still expressed their fearful perception of launching outside of their experience base [PERC]:

- MTI's Roger Boisjoly stressed that "in launching below the data base they were moving away from goodness" [52, p317].
- MTI's Jack Kapp: "Most of the concerns that we had presented were qualitative in nature. At that particular time we had a very difficult time having enough engineering data to quantify the effects that we had been talking about. A lot of it was based on "engineering feel" [52, p308].

Collective testimony coupled with post-event review of the information available to decision-makers shows the use of unsupported judgement regarding risk of cold temperatures to the Challenger vehicle. Without sufficient coherent analysis to substantiate the claim of safety, unsupported judgements were formed about the adequacy of the available technical data and the behaviour of MTI. These were consequentially found to be fallacious, with disastrous consequences.

3.4. Case 3 - UK military aircraft routine structural damage

The UK military aviation regulations require risks to life to be managed within a risk management system, with nominated duty holder personnel being accountable [12]. Authorised engineering management personnel are also permitted to defer maintenance provided it is "considered justifiable and safe" to do so [23, p8]. Deferring maintenance is considered routine since spares,

[16] STS 61-A launched in October 1985 and experienced blow-by and erosion at a calculated O-ring temperature of 75 °F. STS 51-C launched in January 1985 with a calculated O-ring temperature of 53 °F and exhibited the most severe O-ring erosion and blow-by from live mission launches [52, p153].

[17] Launch Commit Criteria were metrics that provided pre-determined thresholds and limitations beyond which a launch could not proceed.

operational considerations, specialist equipment, or expertise are not always available at the point of need for a given maintenance task. Case 3 reviews one such routine request sent to an aircraft airworthiness management team from front-line engineers operating the aircraft. Although it does not end in an accident, it highlights the routine use of unsupported judgement in operational situations.

3.4.1. Limitations of the source

The information was received directly from the engineering staff responsible for the structural airworthiness considerations of the particular military aircraft. Much of the detail is anonymised to protect individuals and teams from scrutiny and to prevent any breach of confidential information regarding the sustainment of military equipment. Although this restricts the level of clarity experienced by the reader, there is sufficient information to identify the use of unsupported thinking in daily aircraft engineering risk decision-making.

3.4.2. Narrative of events

On 3 May 2017, a request was submitted by the operating Squadron to the aircraft airworthiness authority and manufacturer[18] to defer carrying-out a full repair as required by the aircraft maintenance manual. A structural repair was necessary after accidental tooling damage was found on a shaped component that provides the structural connection support between the wing box and the outer wings of the aircraft. The component is known by the manufacturers to be susceptible to fatigue and the damage was spotted during a periodic non-destructive testing (NDT) inspection that preventatively searched for fatigue damage. The operating Squadron were only able to carry-out a partial repair as they were deficient in the necessary tooling to complete the repair at their location. However, as no damage limits for the component were provided in maintenance manuals, a request for airworthiness advice and legal maintenance data was dispatched. The Squadron had no legally approved maintenance data that a partial repair would be sufficient not to cause a structural integrity hazard to the safe operation of the aircraft. A NDT inspection of the repaired area was carried-out as mitigation for the absence of a full repair. The Squadron requested:

"Airworthiness Advice on continued unrestricted Operations... Tooling impact damage to R/H upper [component name ommited for confidentiality]. Pre-blended damage[19] dimensions as follows [dimensions ommitted for brevity]. Damaged areas blended iaw (in accordance with) [structural repair manual reference ommitted for confidentiality] to a ratio of 10-1. NDT [reference

[18] see Footnote 5

[19] Blending is a repair technique that involves the abrasive rubbing and polishing of metal damage in order to smooth out damage such as dents, gouges or scores. Sharp edges create 'stress raisers' in materials under load, and can be initiators of fatigue or overload damage mechanisms.

ommitted] carried out with no fault indications. Roto peening[20] *not carried out due to lack of tooling in theatre*[21] *. Post blend dimensions as follows [dimensions ommitted for brevity].”*

While awaiting a response, the Squadron deferred a full repair and continued operating the aircraft unrestricted. A month later (5 June 2017), approval for the partial repair was received from the manufacturer, via the airworthiness management team. In the interim, the acceptability of the risk to the aircraft had been based on the judgement of the front-line maintenance decision-maker without analytical or deterministic confirmation. The judgement had been carried-out in an unsupported manner.

3.4.3. Unsupported judgement to defer maintenance

Deficiencies in equipment, spares or information to complete engineering activity are frequent in military aircraft maintenance. While designers foresee problems as far as possible and maintainers plan equipment usage and manpower schedules as thoroughly as possible, aircraft operations inevitably generate unexpected damage as a result of human and environmental influences. Deferring maintenance at the point of need allows maintenance managers to use their judgement to balance safety and equipment availability, catering for real-world variations in situation while awaiting manufacturer advice. However, without access to the tools or cognitive ability to generate analytical structural integrity risk assessments, maintainers must proceed non-analytically.

Aircraft maintenance is typically documented scrupulously, recording references to procedures, policies and information used, providing “an audit trail of the work to enable quality assurance, data exploitation and investigations” [23]. The omission of references to quantifiable data, relevant historical events, or analytical assessment supporting the deferment request in section 3.4.2 is conspicuous in its absence. If analytical tools had been used, they would almost certainly have been referenced in this maintenance paperwork entry. This suggests that an unsupported judgement regarding the structural risk was made by the engineering manager, given their personal state of knowledge regarding factors such as: 1) the assumed expected use of the aircraft, 2) the available knowledge and experience of the damage characteristics, 3) that the NDT inspection was suitable and accurate, 4) no other assumed factors would interact with the damage, 5) the impact on remaining structure if the damaged component was to fail. The factors that influence damage propagation are numerous and situationally specific. Decision choices also vary between decision-makers - another engineer faced with the same situation but with

[20] Roto peening is a specialist maintenance activity that imparts a residual compressive load into the surface of metal structure in order to provide tensile stress relief. The process improves resilience to fatigue and some corrosion damage processes.

[21] ‘Theatre’ refers to a deployed military operational location overseas.

a different knowledge base may elect to ground the aircraft while awaiting specialist advice. Descriptive decision analysis provides insight into the motivational and cognitive biases that may cause this (see for example [42], [37], or [36]). Although perhaps counter-intuitively, grounding an aircraft requires just as strong an argument as a decision to accept a fault and continue flying. Grounding aircraft impacts operational availability, may cause reputational harm (depending on the service the aircraft is providing), and commits organisational resources to rectifying the fault. In this case, unsupported judgement of risk posed by the partial repair did not exceed any personal thresholds for intervention. However, if the risk had been deemed unacceptable, the considerations influencing the operational decision-maker may have been:

- Mitigation being considered inadequate, for example the NDT technique might provide insufficient assurance or be considered unreliable [UNTHI].
- Previous experience of this type of repair technique might bias the judgement toward the success or failure of the experience base. This might be considered as a learned association [LA] or an assumption that the repairs are sufficiently similar to draw an inference between the damage situations ([BGKN] and [NBPA]).
- Other personnel may offer up their own experiences, which the manager will weight given their knowledge of the advice-giver, relying on social factors such as reputation and trust [LA] .

Such influences are very personal and impact people to varying degrees. In the case above, the operational urgency and perception of the risk meant that the maintenance manager resorted to their judgement and perception about the risk [PERC], exercising their authority to continue operating while awaiting supportive information.

4. Discussion

Risk assessment provides decision-makers with information about uncertain outcomes in order to help them decide upon the most appropriate action. Supported analytical assessments are desirable from a managerial perspective because they enable normative decision support approaches to be used to justify resource allocation. However, operational conditions may restrict decision-makers to using holistic or unsupported deliberative thinking to construct an assessment of a risky prospect. Reliance by decision-makers on unsupported thinking is not regulated or operationalised, resulting in an opaque decision process, which reduces the level of transparency and accountability that are essential to safety culture-focussed communities. Work by Aven [16] makes progress in this regard by drawing attention to the influence of such thinking on risk scenarios. His method does not benefit decision-makers in operational situations who are unlikely to have the time or means to access a council of experts and risk analysts (as required in Aven [16]) to support the

construction of an unsupported assessment. Better guidance is required for decision-makers in such positions. The following discussion points from the case studies justify this argument.

Type 1 or Type 2 distinction. The decisions made in the cases studied are a result of both Type 1 and Type 2 processes. The influence of Type 1 has bearing on the deliberative Type 2 judgement applied, influenced by emotions such as dread, fear, or a gut feeling of 'rightness', and the time between consideration and action. It is not possible to indicate whether decisions originated from Type 1 intuition and were then justified by Type 2 reasoning, or if the decisions only emerged after Type 2 deliberation. This matters not since the main consideration is to be aware of the extent of unsupported influences on decisions (deliberative, intuitive or otherwise) not their cognitive classification. Had there been better awareness, then the use of unsupported thinking to delay scour reinforcement works (Case 1), or to opt to launch despite conflicting analytical data and judgements that opposed launch (Case 2), may have been given greater scrutiny at the time.

Risk management system compatibility. The risk decisions were made under situational constraints that were incompatible with the risk management processes relevant to each case, compelling decision-makers to use unaided judgement in their decision. The initial decision by the Lamington viaduct track maintenance team caused them to react with an assumption regarding track deformation without knowledge of the known scour risk; the MTI engineers could not adequately quantify their concerns regarding the forecast ambient conditions but intuitively felt the risk was too high; the aircraft Squadron engineering decision-maker made an unaided judgement that the partial repair would not lead to structural collapse.

In the case of Challenger, the decisions, justifications and testimonies indicate that the risk management process was unable to accommodate the holistic and unsupported judgement that prompted numerous key personnel to initially argue a 'no launch' rationale in spite of the technical data available. This observation is not just because the correct decision (in hindsight) had not been selected, but because the risk management system was not configured to give weight to unsupported assessments in operational circumstances [53, 148]. Within NASA's existing technical culture and processes, risk assessments followed scientific and rule-bound engineering standards, but Thiokol's argument conflicted these:

"...observational data, backed by an intuitive argument, were behind all engineering analysis. But subjective, intuitive arguments required lab work and tests before they were considered admissible evidence in FRR (Flight Readiness Review)." [52, p353-354]

In all cases, the holistic representations of risk perceived by individual engineers provided an actionable view on the risk situation but in Cases 1 and 2, the qualitative assessments could not reduce the risk exposure sufficiently. Furthermore, in Cases 1 and 2 the hazards and risks were known to the organisations. They were not black swan events that surprised stakeholders once they emerged, but the uncertainty surrounding risk propagation created sufficient room for arguments both in favour and against intervention. Unable to argue convincingly with holistic risk representations to gain support for timely intervention, the risks were unable to be prevented from becoming manifest.

Triggering more detailed investigation. Unsupported judgement is used in Cases 1 and 3 to guide decision-makers toward escalating intervention and commitment of resources, though stopping short of a slower supported Type 2 assessment.

In Case 3's context, new faults are only referred to the duty holder if the risk is considered (by the maintenance decision-maker) to be a *risk-to-life*. The decision to refer requires either a rule-based trigger or relies on intuition; referring every occurrence as a risky prospect is conservative and unacceptably resource consuming. Regulation [12] provides such a rule-based threshold to the UK military aviation community via the use of risk matrices, but this method is incompatible with structural risks that are difficult to quantify under operational constraints. Furthermore, in acknowledging that quantified risk assessment is a professional judgement that is mainly Type 2 but contains elements of Type 1 [16], maintains that Type 2 reasoning (supported or unsupported) is not judgement-free. As a result, subjective risk matrix output is difficult to validate and of an unknown accuracy owing to insufficient data [13].

In the case of Challenger, unsupported thinking was insufficient to initiate intervention. A number of real-time qualitative cues consistent with the codes in Table 3 may have alerted a need for more robust action. Firstly, to request a launch delay was unprecedented for MTI [52, p305], a clear signal that there was an overwhelming feeling of dread about the hazard [PERC], what Payne and Bettman [40, p123] term problem-focussed coping - treating negative emotion as a signal of decision importance. Secondly, that the normalised 'acceptable' erosion to the O-ring seals was overcome by the single cue that, for the complete Shuttle vehicle system, the launch was outside of any data base for acceptable damage, indicating a conflict with established assumptions [BGKN]. Thirdly, the effect of the cold on the complete system at launch was uncertain and the subscale test data only provided a partially complete representation of the expected vehicle performance ([UNHI] and [BGKN]). The unsupported rationale from MTI's (largely qualitative) rationale [52, p308] that was based on conflicting data was a weak signal for NASA and Marshall engineers. The lack of coherent scientific data enabled differences of opinion on the true impact of the cold effects to be argued, framed by a risk management system that necessitated objective scientific evidence to support risk assessments.

The cases demonstrate that reliance on unsupported thinking for risk decision-making is a consequence of being unable to conduct a methodical, analytical modelling of the scenario. All of the cases indicated that this was owing to operational constraints of the time, resource or cognition available within the decision horizon. This can result in behavioural inconsistencies when referring risks for more detailed appraisal and consideration.

Justifying the use of judgement. In Case 2, the MTI team had incomplete data and were ineffective in justifying their claims. Their holistic argument that the ambient temperature conflicted with the database for the complete vehicle was weaker than the competing rationally-presented argument concerning other available data. Had the advocates of 'no launch' been able to present their argument more coherently, they may have gained more support. Justification is also lacking in Case 3, where the engineering manager had not explained the grounds for deferral. The Toulmin [54] theoretical framework for presenting an argument provides a generalised logical structure for the information being used to make a claim. Briefly, it states that in order for a conclusion to be drawn from the foundation of an argument (the data), there must be a bridging inference-license, rule or principle that exists to justify the step from data to conclusion. In Case 3, such a logical bridge is not apparent between the available information (damage dimensions, the non-destructive test for subsurface damage indications and the use of a recognised repair technique) and the claim that deferral was appropriate. Although links can be inferred, the perception of acceptability depends on the knowledge base of the inferencer. For instance, where the deferral wording reads "NDT carried out with no fault indications", there is an assumption that the NDT was 100% accurate, which may not be true and can be open to challenge. Improvements in the ability to justify holistic, non-analytical arguments in a logical but situationally-specific rational manner, particularly in the face of hard data, will be crucial to incorporating such thinking into risk assessment.

4.1. Future development opportunities

Drawing attention to the use of unsupported judgement is useful, but converting this attention into a practical tool requires further work. By considering decision-maker rationality as situationally-specific, a bounded view of the context in which a decision is being made could help decision-makers form an "ecologically rational"[22] decision basis [55]. This may be done by applying defined, situationally validated heuristic processes, such as Fast and Frugal Trees. In some situations, these methods have provided comparable performance to more complicated methods requiring greater understanding such as linear and logistic regression ([56–58]). However, handling engineering decision situations by fast and frugal means steps boldly away from the

[22]The term ecological rationality refers to funtional matches between cognition and environment [55, p1671].

formalised mechanisms, models and analytical devices that provide more traditional forms of decision support. While not denigrating or diminishing the role that such methods play in supporting decision-makers, there is a need to explore simpler, faster methods which can still provide some level of support in highly constrained, operational situations.

Another consideration for tool development is the influence of wider system elements on risk, such as management policies and human factors. “In complex systems, major failures seldom result from a single problem, but are typically caused by multiple problems that coincide to allow catastrophic failure” [59]. The SAM framework has been used to model how management decisions influence engineering risk [60]. Although, modelling behaviour is not immediately useful for operational maintenance decision-makers, the concept demonstrates how risk can be linked to behaviour and organisational policies through an influence diagram.

Justifying arguments logically [54] can provide a useful foundation in operational engineering decision-making, having already found utility in engineering design [61]. Being able to document unsupported decision justifications would enable post hoc learning in decision-making, and benefit risk communication. Both aspects are also crucial tenets within risk governance, an emerging framework that advocates new ways of conceptualising complex, uncertain and ambiguous tasks [45]. Externalising one’s own argument may also help individual decision-makers in situations with no access to a committee (as required by Aven [16]), and provide a much needed audit trail for regulated industries that necessitate this.

5. Conclusion

This paper has sought to highlight that greater attention should be paid to how risk decision-makers are influenced by operational conditions that restrict resources, information and time. The case studies show that in resource-limited, information-poor or time-constrained situations, deliberative judgements are made but are not always supportable by analytical information within the decision horizon. Within organisational risk management constructs, this invites criticism of why judgement and holistic perspectives have not been more openly accounted for in risk processes. Addressing naturalistic risk decision-making and judgement use contributes to the safety domain’s growing momentum toward improving system safety resilience, handling unforeseen events [19] and reality-based safety science [**?**].

We have suggested that the decision hierarchy of Howard and Abbas [24] should be extended, particularly to address operationally constrained decision situations with insufficient time available to reach a requisite decision (Figure 1). We have also argued that a previous effort to relate dual process theory to risk analysis [16] was flawed, and that dependence upon a committee-based

solution could not practically support the kind of real-world operational maintenance situations presented here. When combined with the case study findings, we conclude that the additional distinction suggested here of unsupported/supported Type 2 thinking to the dual process theory (Figure 2) is useful in risk decision-making. Highlighting the influence of physical factors on unsupported Type 2 risk decision-making could provide helpful risk management information.

Forming decisions from unsupported thinking is hazardous because it is even more susceptible to cognitive and motivational biases than supported thinking. Without a committee to debias judgements [16] or provide consensus, alternative approaches are necessary to support operational maintenance decision-makers. Some promising avenues of work have been highlighted, including the use of argument justification and prescribed situationally-specific heuristics. The utility of such approaches requires testing with real-world decision-makers to prove the credibility and suitability for their use when handling risky prospects in pressured naturalistic operational situations.

Addressing unsupported thinking in risky, operational situations may support decision-makers in applying their own judgement, improve the transparency of decision-making and provide better understanding of what information was available at the time. This should also enable organisations to better exploit corporate learning from such situations.

References

[1] J. T. Luxhøj, M. Morton, Probabilistic Safety Risk Analysis in Complex Domains : Application to Unmanned Aircraft Systems, in: 11th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, September, ISBN 978-1-60086-941-9, 1–13, doi:10.2514/6.2011-6920, 2011.

[2] T. Uyar, Structuring Risk Assessment Process with Tallying in Aviation Safety Management, The International Journal of Aerospace Psychology 0 (0) (2019) 1–9, ISSN 2472-1840, doi:10.1080/24721840.2019.1621176.

[3] L. Chen, E. Arzaghi, M. M. Abaei, V. Garaniya, R. Abbassi, Condition monitoring of subsea pipelines considering stress observation and structural deterioration, Journal of Loss Prevention in the Process Industries 51 (December 2017) (2018) 178–185, ISSN 09504230, doi:10.1016/j.jlp.2017.12.006.

[4] I. Animah, M. Shafiee, Condition assessment, remaining useful life prediction and life extension decision making for offshore oil and gas assets, Journal of Loss Prevention in the Process Industries 53 (2018) 17–28, ISSN 09504230, doi:10.1016/j.jlp.2017.04.030.

[5] M. Vagnoli, R. Remenyte-Prescott, J. Andrews, Railway bridge structural health monitoring and fault detection: State-of-the-art methods and future challenges, Structural Health Monitoring 17 (4) (2018) 971–1007, ISSN 17413168, doi:10.1177/1475921717721137.

[6] M. I. Rafiq, M. K. Chryssanthopoulos, S. Sathananthan, Bridge condition modelling and prediction using dynamic Bayesian belief networks, Structure and Infrastructure Engineering 11 (1) (2015) 38–50, ISSN 17448980, doi:10.1080/15732479.2013.879319.

[7] M. Gobbato, J. P. Conte, J. B. Kosmatka, C. R. Farrar, A reliability-based framework for fatigue damage prognosis of composite aircraft structures, Probabilistic Engineering Mechanics 29 (2012) 176–188, ISSN 02668920, doi:10.1016/j.probengmech.2011.11.004.

[8] F. Khan, S. Rathnayaka, S. Ahmed, Methods and models in process safety and risk management: Past, present and future, Process Safety and Environmental Protection 98 (2015) 116–147, doi:10.1016/j.psep.2015.07.005.

[9] W. Rogers, Space Shuttle Challenger Accident Commission Report URL `https://spaceflight.nasa.gov/outreach/SignificantIncidents/assets/rogers_commission_report.pdf`.

[10] C. Haddon-Cave, The Nimrod Review: An independent review into the broader issues surrounding the loss of rhe RAF Nimrod MR2 Aircraft XV230 In Afghanistan in 2006, Tech. Rep., The Stationary Office, London, URL `https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/229037/1025.pdf`, 2009.

[11] Society for Risk Analysis, Society for Risk Analysis Glossary, URL `https://www.sra.org/resources`, 2018.

[12] Military Aviation Authority, Regulatory Article 1210: Ownership and Management of Operating Risk (Risk to Life), 5, UK Ministry of Defence, URL `https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/760768/RA1210_Issue_5.pdf`, 2014.

[13] L. A. Cox, What's Wrong with Risk Matrices?, Risk Analysis 28 (2) (2008) 497–512, ISSN 02724332, doi:10.1111/j.1539-6924.2008.01030.x.

[14] F. Goerlandt, N. Khakzad, G. Reniers, Validity and validation of safety-related quantitative risk analysis: A review, Safety Science 99 (2017) 127–139, doi:10.1016/j.ssci.2016.08.023.

[15] E. Zio, The future of risk assessment, Reliability Engineering and System Safety 177 (June 2017) (2018) 176–190, ISSN 09518320, doi:10.1016/j.ress.2018.04.020.

[16] T. Aven, How the integration of System 1-System 2 thinking and recent risk perspectives can improve risk assessment and management, Reliability Engineering & System Safety 180 (February 2017) (2018) 237–244, ISSN 09518320, doi:10.1016/j.ress.2018.07.031.

[17] J. S. B. T. Evans, Reflections on reflection: the nature and function of type 2 processes in dual-process theories of reasoning, Thinking & Reasoning 25 (4) (2019) 383–415, ISSN 1354-6783, doi:10.1080/13546783.2019.1623071.

[18] A. Rae, D. Provan, H. Aboelssaad, R. Alexander, A manifesto for Reality-based Safety Science, Safety Science 126 (September 2019) (2020) 104654, ISSN 18791042, doi:10.1016/j.ssci.2020.104654, URL https://doi.org/10.1016/j.ssci.2020.104654.

[19] P. Swuste, J. Groeneweg, C. van Gulijk, W. Zwaard, S. Lemkowitz, Y. Oostendorp, The future of safety science, Safety Science 125 (September 2019) (2020) 104593, ISSN 18791042, doi:10.1016/j.ssci.2019.104593, URL https://doi.org/10.1016/j.ssci.2019.104593.

[20] X. Yang, S. Haugen, Classification of risk to support decision-making in hazardous processes, Safety Science 80 (2015) 115–126, ISSN 09257535, doi:10.1016/j.ssci.2015.07.011.

[21] European Aviation Safety Agency, Easy Access Rules for Airworthiness and Environmental Certification (Regulation (EU) No 748/2012), URL https://www.easa.europa.eu/sites/default/files/dfu/Easy_Access_Rules_for_Airworthiness_and_Environmental_Certification.pdf, 2018.

[22] Public Records Act 1958 (6, 7 Eliz.2), URL http://www.legislation.gov.uk/ukpga/Eliz2/6-7/51, 1958.

[23] Military Aviation Authority, Regulatory Article 4813: Maintenance Records (MRP145.A.55), in: 4000 series: Continuing Airworthiness Engineering Regulations, UK Ministry of Defence, 4, URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/760707/RA4813_Issue_5.pdf, 2018.

[24] R. A. Howard, A. E. Abbas, Foundations of Decision Analysis (Global Edition), Pearson Education, Harlow, first edn., ISBN 9781292079691, 2016.

[25] S. M. Tavares, P. M. de Castro, An overview of fatigue in aircraft structures, Fatigue and Fracture of Engineering Materials and Structures 40 (10) (2017) 1510–1529, doi:10.1111/ffe.12631.

[26] T. Theofanous, Risk Assessment and Management, in: I. Milne, R. Ritchie, B. Karihaloo (Eds.), Comprehensive Structural Integrity, chap. 1.19, Elsevier, Oxford, 1st edn., 2003.

[27] J. L. Hurtado, P. Hoffman, Airframe Integrity Based on Bayesian Approach, in: Proceedings - Annual Reliability and Maintainability Symposium, vol. 1677444, ISBN 1424400082, 630–635, 2006.

[28] D. Cope, J. Moffett, Bayesian Updating of Structural Health and Integrity Assessments using Real-Time Inspection Results, in: 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference; 20th AIAA/ASME/AHS Adaptive Structures Conference; 14th AIAA, April, American Institute of Aeronautics and Astronautics, Reston, Virigina, ISBN 978-1-60086-937-2, 1–19, doi:10.2514/6.2012-1438, 2012.

[29] D. Rusk, K. Lin, D. Swartz, G. Ridgeway, Bayesian updating of damage size probabilities for aircraft structural life-cycle management, 19th AIAA Applied Aerodynamics Conference (c), doi:10.2514/6.2001-1646.

[30] R. F. Torregosa, W. Hu, Probabilistic risk analysis of fracture of aircraft structures using a Bayesian approach to update the distribution of the equivalent initial flaw sizes, Fatigue and Fracture of Engineering Materials and Structures 36 (11) (2013) 1092–1101, doi:10.1111/ffe.12103.

[31] J. Yang, S. Manning, Aircraft Fleet Maintenance Based on Structural Reliability Analysis, Journal of Aircraft 31 (2) (1994) 419–425.

[32] D. Straub, Stochastic Modeling of Deterioration Processes through Dynamic Bayesian Networks, Journal of Engineering Mechanics 135 (10) (2009) 1089–1099, ISSN 0733-9399, doi:10.1061/(ASCE)EM.1943-7889.0000024.

[33] D. Straub, A. Der Kiureghian, Bayesian Network Enhanced with Structural Reliability Methods: Methodology, Journal of Engineering Mechanics 136 (10) (2010) 1248–1258, ISSN 0733-9399, doi:10.1061/(ASCE)EM.1943-7889.0000173.

[34] D. Straub, I. Papaioannou, Bayesian Updating with Structural Reliability Methods, Journal of Engineering Mechanics 141 (3) (2015) 04014134, ISSN 0733-9399, doi:10.1061/(ASCE)EM.1943-7889.0000839.

[35] A. Rae, R. Alexander, Forecasts or fortune-telling: When are expert judgements of safety risk valid?, Safety Science 99 (2017) 156–165, ISSN 18791042, doi:10.1016/j.ssci.2017.02.018.

[36] P. Baybutt, The validity of engineering judgment and expert opinion in hazard and risk analysis: The influence of cognitive biases, Process Safety Progress 37 (2) (2018) 205–210, ISSN 10668527, doi:10.1002/prs.11906.

[37] C. L. Brown, D. R. Utley, Ambiguity Aversion in Engineers, Engineering Management Journal 31 (1) (2019) 2–7, ISSN 1042-9247, doi:10.1080/10429247.2018.1503037.

[38] Military Aviation Authority, Regulatory Article 4812 - Certification of Air System Release and Component (MRP 145.A.50), in: 4000 series: Continuing Airworthiness Engineering Regulations, 5, UK Ministry of Defence, 1–10, URL https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/760711/RA4812_Issue_5.pdf, 2018.

[39] P. Slovic, Finucane,Melissa L., E. Peters, MacGregor,Donald G., Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality, Risk Analysis 24 (2) (2004) 311–322, doi:10.1111/j.0272-4332.2004.00433.x.

[40] J. W. Payne, J. R. Bettman, Walking with the Scarecrow: The Information-processing Approach to Decision Making, in: D. J. Koehler, N. Harvey (Eds.), Blackwell Handbook of Judgment and Decision Making, chap. 6, Blackwell, Oxford, second edn., ISBN 1-4051-0746-4, 110–132, 2005.

[41] D. Over, Rationality and the Normative/Descriptive Distinction, in: D. J. Koehler, N. Harvey (Eds.), Blackwell Handbook of Judgment and Decision Making, chap. 1, Blackwell, Oxford, second edn., ISBN 1-4051-0746-4, 4–18, 2005.

[42] A. Tversky, D. Kahneman, Judgment under Uncertainty: Heuristics and Biases, Science 185 (4157) (1974) 1124–1131, ISSN 0036-8075, doi:10.1126/science.185.4157.1124.

[43] D. Kahneman, A. Tversky, Prospect Theory: An Analysis of Decision under Risk, Econometrica 47 (2) (1979) 263–291, URL http://www.jstor.org/stable/1914185.

[44] G. Keren, K. H. Teigen, Yet Another Look at the Heuristics and Biases Approach, in: D. J. Koehler, N. Harvey (Eds.), Blackwell Handbook of Judgment and Decision Making, chap. 5, Blackwell, Oxford, second edn., ISBN 1-4051-0746-4, 89–109, 2005.

[45] M. B. van Asselt, O. Renn, Risk governance, Journal of Risk Research 14 (4) (2011) 431–449, ISSN 1366-9877, doi:10.1080/13669877.2011.553730.

[46] HSE, Reducing Risks, Protecting People, HSE Books, Norwich, first edn., ISBN 0-7176-2151-0, URL https://www.hse.gov.uk/risk/theory/r2p2.pdf, 2001.

[47] S. O. Hansson, T. Aven, Is Risk Analysis Scientific?, Risk Analysis 34 (7) (2014) 1173–1183, ISSN 02724332, doi:10.1111/risa.12230.

[48] N. Pidgeon, Risk assessment, risk values and the social science programme: why we do need risk perception research, Reliability Engineering & System Safety 59 (1) (1998) 5–15, ISSN 09518320, doi:10.1016/S0951-8320(97)00114-2.

[49] M. Saunders, P. Lewis, A. Thornhill, Research Methods for Business Students, Pearson Education, Harlow, seventh edn., ISBN 978-1-292-01662-7, 2016.

[50] Rail Accident Investigation Branch, Rail Accident Report - Structural failure caused by scour at Lamington viaduct, South Lanarkshire 31 December 2015, Tech. Rep., Rail Accident Investigation Branch, URL https://www.gov.uk/raib-reports/structural-failure-at-lamington-viaduct, 2016.

[51] T. Aven, Risk assessment and risk management: Review of recent advances on their foundation, European Journal of Operational Research 253 (1) (2016) 1–13, ISSN 03772217, doi:10.1016/j.ejor.2015.12.023.

[52] D. Vaughan, The Challenger Launch Decision - Risky Technology, Culture, and Deviance at NASA, The University of Chicago Press, Chicago, ISBN 0226851753, 1996.

[53] U. Hoffrage, J. N. Marewski, Unveiling the Lady in Black: Modeling and aiding intuition, Journal of Applied Research in Memory and Cognition 4 (3) (2015) 145–163, ISSN 22113681, doi:10.1016/j.jarmac.2015.08.001.

[54] S. Toulmin, The Uses of Argument, Cambridge University Press, Cambridge, 1 edn., ISBN 521092302, 1958.

[55] S. Mousavi, G. Gigerenzer, Risk, uncertainty, and heuristics, Journal of Business Research 67 (8) (2014) 1671–1678, ISSN 01482963, doi:10.1016/j.jbusres.2014.02.013.

[56] G. Gigerenzer, W. Gaissmaier, Heuristic Decision Making, Annual Review of Psychology 62 (1) (2011) 451–482, ISSN 0066-4308, doi:10.1146/annurev-psych-120709-145346.

[57] S. Hafenbrädl, D. Waeger, J. N. Marewski, G. Gigerenzer, Applied Decision Making With Fast-and-Frugal Heuristics, Journal of Applied Research in Memory and Cognition 5 (2) (2016) 215–231, ISSN 22113681, doi:10.1016/j.jarmac.2016.04.011.

[58] L. J. Gibbons, K. Stoddart, Fast and frugal heuristics': Clinical decision making in the Emergency Department, International Emergency Nursing 41 (April) (2018) 7–12, ISSN 1755599X, doi:10.1016/j.ienj.2018.04.002.

[59] D. M. Murphy, M. E. Pate-Cornell, The SAM Framework: Modeling the Effects of Management Factors on Human Behavior in Risk Analysis, Risk Analysis 16 (4) (1996) 501–515, ISSN 0272-4332, doi:10.1111/j.1539-6924.1996.tb01096.x.

[60] M. E. Pate-Cornell, The Engineering Risk Analysis Method and Some Applications, in: W. Edwards, R. F. Miles, D. von Winterfeldt (Eds.), Advances in Decision Analysis, chap. 16, Cambridge University Press, Cambridge, first edn., ISBN 9780521682305, 302–324, 2007.

[61] T. Polacsek, S. Sharma, C. Cuiller, V. Tuloup, The need of diagrams based on Toulmin schema application: an aeronautical case study, EURO Journal on Decision Processes 6 (3-4) (2018) 257–282, ISSN 2193-9438, doi:10.1007/s40070-018-0086-3.

Appendix G

Thematic analysis coding and sorting photographs

The figures provided illustrate the sorting of codes identified during the Phase 2 survey results analysis, and arrangement in to themes. Codes were initially captured on to an individual note card for each code/phrase (Figure G.1). Codes were then grouped together, and a different coloured note card was detailed with the connecting theme (Figure G.2).

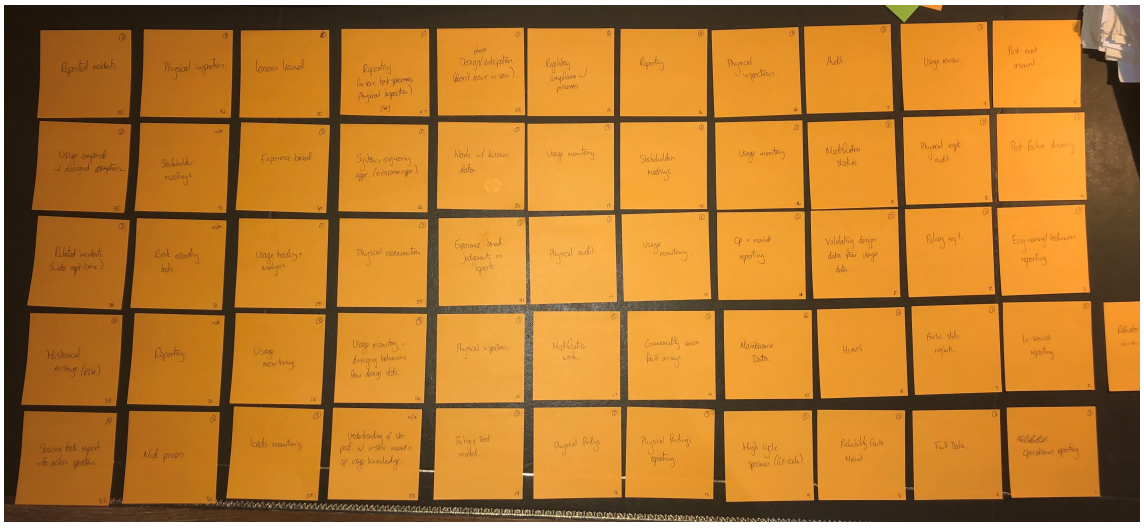


Figure G.1: Survey results thematic analysis typical collection of themes, elements or factors prior to sorting and grouping.

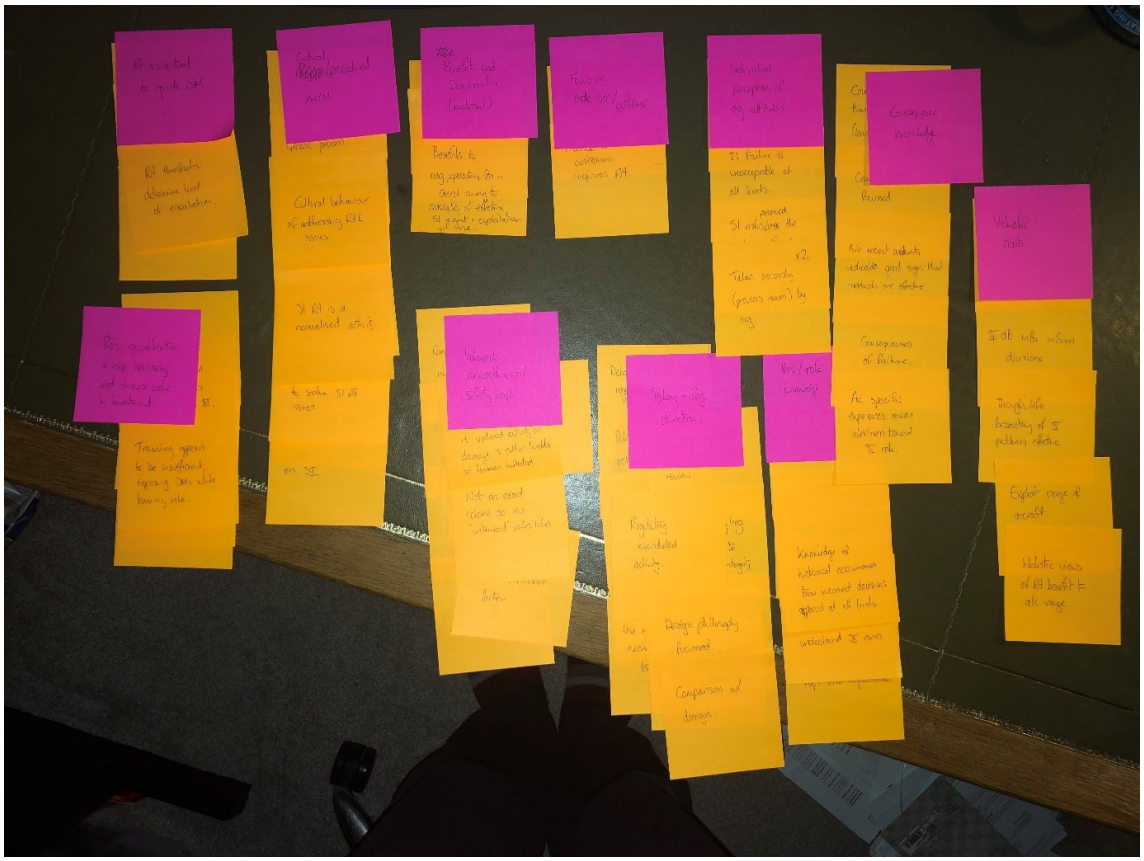


Figure G.2: Survey results thematic analysis typical collection of themes following grouping into codes.

Appendix H

Outstanding survey responses

Several survey questions were not included in the thesis main body, as they did not contribute or amplify the points already made. However, for completeness, they are included in this appendix.

Question 27 - "Do you have any comments on the answers provided in this section?" As an open question, a variety of responses were received, although there were too few related comments to create a thematic map. Responses that are related are grouped below:

Assumptions and uncertainties The importance of assumptions and uncertainties is remarked in the two responses received. The two comments infer that assumptions are dependent upon the knowledge of individuals, which for some decision-makers unfamiliar with engineering contexts may result in a lapse in understanding.

Regulatory/policy maker *"Duty Holders understand risk 'uncertainty' far better than 'engineering assumptions', simply because they deal in uncertainty daily in all of their RtL, but only rarely engineering assumptions."*

Regulatory/policy maker *"Assumptions and uncertainties are at the heart of SI risk management and my career has shown that many near-misses have been resulted from good fortune rather than expert predictions. The trick is to weave that experience into the bounding of assumptions and recognising the associated uncertainties so that the resulting judgements are justified."*

Structural health monitoring One participant identified with SHM providing decision criteria, which relates the discussion about the prospects for SHM in Section2.2.1.1:

Operator/end user "The SHM system is designed to provide go/no-go outputs which does not provide much information to allow considered risk decisions."

Types of risk assessment One interview participant remarked about two types of risk assessment. Risks that impact individual aircraft are treated differently to fleet-wide issues:

Operator/end user *"Only that I think there's two types of risk assessment that we're talking about. There's kind of the operational "I've got a decision on one aircraft that I need to make a decision about" and "actually I'm make a decision for the fleet" and where we've been discussing some of the uncertainties or assumptions you've got more, it feels like you've got more time in those fleet decisions to probe that little bit further, so you're less likely to have those issues I think when you're looking at fleet wide bits because it tends to be that those longer term problems, you'll not only do over the phone, email and tech queries, but you'll probably end up being discussing those issues in person in either formal meetings and also in the sidelines of meetings as well, so I think that reduces the risk so, there's a lot to be said for those structured meetings."*

Other comments Comments that were not pertinent to topical aspects were remarked as follows:

Regulatory/policy maker *"I am too remote from risk-based decision making, and/or it is so rare relating to SI, for me to provide examples."*

Regulatory/policy maker *"Sorry, Rich, this survey is too long to answer every question in the time I have available"*

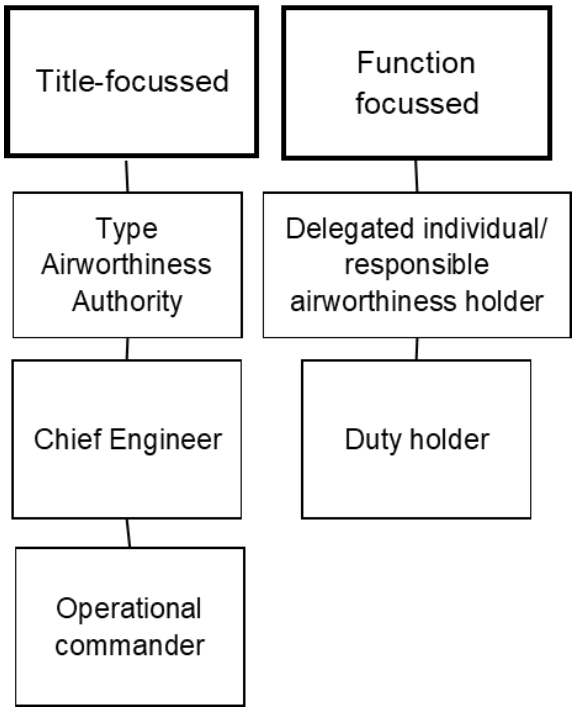


Figure H.1: Survey Questions 29b and 29c thematic analysis results.

Question 29b - "If you were/are not able to decide upon a particular in-service structural integrity risk, who does make the ultimate decision?" and Question 29c - "Who is the ultimate Decision Maker regarding in-service structural integrity risks?" The responses to these two questions provide the same output, but were arrived at through different responses to Question 29[1]. The responses were combined and a thematic analysis established the simple thematic map in Figure H.1. This illustrates that respondents generally defer to higher authorities, named by either role title or function. The identification of both duty holding and airworthiness authorities as the ultimate decision-maker suggests that escalatory routes may depend on situations or organisational constructs.

Question 37 - "Do you have any comments on the answers provided in this section?" The only comments that were connected regarded feedback about the survey:

[1] Question 29 - "Were you previously, or have you currently been formally delegated the authority to own and decide upon structural integrity risks in your position?"

Survey feedback Two participants that were interviewed provided feedback on the questions:

Operator/end user *"Just that the questions were excellent."*

Operator/end user *"So I mean in terms of how I answered, it's actually quite difficult really, I felt unsure of how to answer a lot of those questions...Sometimes it was understanding the question, but even when I understood it, it was, oh that's kind of put me on the spot a little bit, it's given me a little bit of an eye opener, and well the ASIC was what 6 years ago I did that, probably the last formal training I had in structural health was the ASIC and maybe the RCM that I did, other than that it's just been on the job-type training."*

Structural health monitoring One respondent commented on the status of structural health monitoring and it's influence on commercial aspects of civilian aircraft operations:

Designer/equipment manufacturer *"For health monitoring. It has been studied, but for commercial A/C if a damage is known during the operation phase, there is a duty to repair, leading to commercially non-viable solution."*

Unknown unknowns One respondent remarked that regulatory frameworks restrict creative thinking about unknown unknown threats, which is mitigated in UK military aviation by the use of independent oversight:

Regulatory/policy maker *"Black Swans are easy to see in hindsight but, as mentioned, the regulatory framework does not facilitate holistic assessment of risk that might otherwise spot them. This is a particularly valuable aspect of the inclusion of independent oversight of the SI risk management processes employed in UK military aviation."*

Appendix I

Framework development requirements

I.1 User requirements

Table I.1: User requirements generated from the Phase 2 survey results.

Survey section	Survey question	Requirement
1	1, 2, 5	Methodology applicability to aviation, operator/end users, risk decision makers and risk analysts
1	4	Support risk analysis of structural integrity issues specifically
2	6	Define and follow recognisable process: Risk identification; Treatment; Managed.
2	6	Generate a threshold or subjective argument for: When external support is necessary; When escalation of the decision is required
2	6	Account for individualistic aspects: risk perception; risk appetite
2	6, 9	Allow users to link directly between the cues applied and information available at the time of the decision, to the decision choice made: Technical information, fault information, fault history and holistic applicability; Policy-directed constraints (including thresholds [Q9]) or support for decision basis; Decision alternatives; Operational need and situational constraints requiring judgement to be used; Define the expected consequences and the robustness, assumptions or uncertainties of the decision relative to the consequences defined.

Table I.1: (continued)

Survey section	Survey question	Requirement
2	6	Method is consistent from ‘identification’ to ‘managed’: Justify why the risk is identified; justify how the risk should be treated; justify when an issue is considered managed.
2	7	Method to be teachable and easy to revisit once outcomes are known: Improve the learning of decision situations; Vicarious learning from peer experiences, the situational specifics and information available to support the decision.
2	7	Conceptual and material benefit from the method to the users is necessary to support belief in the method
2	8	Method able to be populated by both numeric and non-numeric risk values: Mode (qual/quan) selected should reflect the quality of the information available to the analyst and decision-maker; “poor” information settings should be justified by verbal at the very least; enable input from other risk analysis methods is appropriate (e.g. if QRA is complete, or Bow Tie informs of pre-established barrier information)
2	9	Method should provide a verbalised threshold indication, to indicate the defining aspect for intervention being taken or not.
2	9d1a	Method must be simple to suit the operational environment, but enable inputs from complex methods (completed elsewhere).
2	10	Method must be usable for speculation on potential risks (to support proactive risk decision making).
2	11	Method does not make use of “likelihood” for verbal assessment of risk.
2	13, 14, 15, 16	Enable communication of uncertainties [Q13], assumptions [Q14], the use of conservatism [Q15] and when human subjectivity has been applied [Q16]. Clear flags to these criteria enable the risk assessment to be challenged in the event their conditional validity becomes compromised by new information.

Table I.1: (continued)

Survey section	Survey question	Requirement
2	18	Method should not be reliant on specialist skills, or burdensome in application. Method should be framed by the environmental context and organisational process. Method should be shaped by experience and subjectivity of individual, enabling the development of such.
3	19	Method documents the information necessary for a decision: Environmental/situational specifics; Practical mitigation considerations; Measurable fault information, causal factor and perceived vulnerability; Personal threshold/need for advice/when to act; Risk: identification and consequence (likelihood may not be possible initially).
3	20	Follow a generalised decision process: Gather information (Situation assessment, defining the problem) on fault information and deterioration visualisation, perception of risk, operational requirement and urgency, and decision alternatives available given fault and situation (Generating/considering one and response options); interim action (Generating/considering one and response options) as seek advice, escalate, or improve risk assessment if possible; intervention (Outcome review) including justifying decision basis and taking action.
3	21	Method captures decomposed judgement through input, processing and output aspects if applicable.
3	21	Method enables users to apply judgement and experience to justify a decision or argument.
3	21	Method identifies the shortcomings of the assessment to enable improvement: knowledge (access to specialist analysis, reliance on peer review/corp knowledge); information (relevance of related situations, fault data, component data, info sharing with stakeholders); analysis (QRA generation).
3	22	Method enables justification by being: able to justify individual judgement; written/documented; policy followed; subject to peer review/scrutiny; accounting for stakeholder working groups.
3	23	Method permits both numerical and non-numerical representations of the risk situation.

Table I.1: (continued)

Survey section	Survey question	Requirement
3	24	Enable decision makers to use any risk input as a component of the overall decision basis (being risk-informed).
3	25	Method should be robust to operational pressure - easy to use, recall and quick to apply, easily accessible within the working memory for when multiple conflicting tasks will be fighting for cognitive resource.
3	27	Method should enable clear identification justification of assumptions and uncertainties for the benefit of non-SME decision-makers, even under pressure.
3	27	Method should enable clear justifiable indication of whether other equipment is likely to be effected.
3	27	Method should enable transparency of the decision basis from a decision-maker perspective.
3	27	Method enables a clear link between the risk assessment and the decision-maker realms.
4	30	Method should incorporate methods that provide apparent transparency for users: procedures and tools; involve stakeholders; enable and evidence independent involvement; record risk activity (risk information, record the decision, reporting operator risk activity).
4	31	Method should be targeted primarily towards op/end users.
4	32	Method enables inclusion of black swan events.
4	34	Method provides users with confidence in assessment: uses data is available; applies common organisational guiding principles of process, independence and assurance; people-focussed: peer and SME review, group or one-to-one, stakeholder involvement.
4	35	Method enables audit of policy compliance and assurance irrespective of enduring nature of risk issue.
4	35	Method enables a basic risk refinement step at the earliest stage in order to identify the need for escalation (is the issue a risk requiring further treatment, mitigation, intervention, reporting, involvement of stakeholders, resource allocation).

Table I.1: (continued)

Survey section	Survey question	Requirement
4	36	Method should be common across subjective and non-subjective risk assessments.
4	36	Method should be able to provide audit trail of the information used, decisions and uncertainties, be understandable by non-SMEs, and increase sharing of experiences.

I.2 System requirements trace

System Requirements		*Development Model x-check*	
		v8	**Remarks**
Test Strategy	Target audience to prove new method.	Test schedule	
	Show general behavioural tendencies	Test schedule	
Applicability	Project definition	Context	
	Conditions of use: not strategic decisions or emergency decisions; framed by sociotechnical system.	Y	Primary gate requires consideration of whether the means, advice, context or available knowledge-base compels judgement use. In other words, are there no other means by which a supported judgement can be levied.
Generality	Operational decision-making centric	Y	Designed around a fast and frugal tree with individualistic but unambiguous 'gates' for evaluating decision situations, that require justification to validate the selected option. Method is flexible to the situation and not intended to be dependent upon specific resources being available.
	Flexible to context	Y	Method does not require utilisation of particular tools, but responses depend upon the user's skill level and available resources.
	Handle prospective situations	Y	Method depends upon argument, which can be formed around current, past or future issues and depends only upon the evidence that is used to support the premisses.
	Communicate non-objective factors	Y	Framework instructs users to justify why a proposition is so. If a proposition depends upon individualistic experiences or beliefs necessitates then this must be arranged to logically complete the argument.
	Generalised process: gather info/SA/define the problem - interim action/generating and considering options - intervention/outcome review	Y	Options generation not required as already defined by heuristic termination points.
Holistic Process	Prompt for either holistic risk management or individual treatment	N	Wider risk management space yet to be developed.
	Focus on judgement on risk	Partly	Mental model vs environment: individually developed and applied to a given situation, which will identify the cues acted upon. Framework doesn't direct which mental model to apply. Info required to improve: May be detailed in the rebuttal, but not explicitly detailed as required. **May need to be addressed through training, or an additional prompt in the model.** Learning from models: dependent upon feedback and learning from experienced practitioners and objective evidence. This is achieved by documented the argument components and through personal development and revisiting decisions.
	Trigger for use of framework	Y	Primary gate requires a situation to be recognised as necessitating judgement use before it is permitted to be applied.
	Follow generalised decision process. Support the wider risk and decision-making domain (alignment with literature on process)		
	Structure the model for simplicity and learning (simple, memorable, easy to revisit) - visual, use mnumonic or metaphor, alliteration, etc.	Partly	Built on heuristic foundations. **Metaphor (safe judgement) being explored to make more memorable.**
	Able to be applied by non-specialists in operational contexts.	Y	Does not necessitate specialist skill, flexes to the skill level of the user and rejects the use of judgement if user is unable to soundly generate an argument.
	Enables auditability and transparency of the decision basis (independent understanding, non-SME understanding, revisitable for learning).	Y	Requires documenting of the argument in written form, and direct users to justify the step that concluded their decision making.

System Requirements		*Development Model x-check*	
		v8	**Remarks**
	Enables auditability and transparency of the decision basis (independent understanding, non-SME understanding, revisitable for learning).	Y	Requires documenting of the argument in written form, and direct users to justify the step that concluded their decision making.
Output	Display? (time sensitive information: info available, info acted upon. Qualifiers and rebuttals: assumptions, uncertainties, conservatisms)	Partly	**Opportunity to display is provided by the method, but the defined elements are not explicit in requiring detail by user.** Time sensitive information is inherent in the argument (unless the user is untruthful), and can be ratified by consensus. Ass/uncert etc. are communicated through the use of argument rebuttals.
	Justification: escalation requirements; threshold for action; reasonability of action.	Y	Justification is required after tertiary gate, but "as required" for primary and secondary if the user deems it necessary. The need to justify escalation of a risky decision (which is in itself a decision) may have greater significance for some stakeholders than for others. Threshold for action is determined by the gate at which users reach termination, e.g. proceeding to tertiary gate requires users believe they're able to generate a sound argument. Reasonability of action is arguable by the user through the strength of their argument.
Alternate Choices	Choices	Y	Choices are dictated by the heuristic choices at each gate. Although not objective, choice thresholds at each gate are individualistically driven and provide the framework to justify choice thresholds. Gate outcomes result in unambiguous alternatives being selected.
	Justify	Y	Alternate choices are defined by the decision archetype, specifying either fly or no fly, with a requirement to justify the choice if the situation requires. The argument form can be exploited to describe influences on judgement validity through rebuttals (i.e. factors that, if changed, would alter or reduce the strength of the argument). Assmt of speculative events and disregard of low P events is inherent to any argument, e.g. argue to "fly" because of warrant meaning low P event is disregarded.
	What is the hierarchy of choices?	Partly	Established by the fast and frugal tree in order: need to use judgement (because judgement is bias and less objective meaning errors are not easily attributable); basic test of judgement argument soundness; debias/consensus for judgement. Robustness is established regarding the argument at tertiary stage through debias/consensus but **physical system robustness (e.g. by recognising redundancy in a system or introducing performance limitations) are not explicitly ruled upon by the framework.** This is to enable flexibility .
	Dynamic aspects	Y	Document version control enables the assessment to be revised, **although this is not planned to be tested.**
External support	Justify need for support/escalation	Partly	This features within the alternatives presented, if the user answers "no" at Gate 2 or 3. Although, the **"wide reaching ramifications" perspective is a consideration for the holistic handling of the risk (other aircraft affected)**.
	Use of external assessment - justify use/trust	Y	If external assessment is available, the framework prompts users to consider the argument basis at Gates 1 and 2 (signified by the dotted lines to indicate the need to justify is discretionary and dependent upon the situation).
	Enable inclusion of external support (qualitative or quan)	Y	In addition to enabling users to consider external support as a "supported" assessment and therefore not requiring judgement (Gate 1 decision option), external support may also be considered within the formation of argument to provide warrants or backing.
Cues	Capture individual reaction	Y	The formation of the argument around a number of cues and the perception of their validity is the foundation to justifying a particular decision.
	Structure cues, label, and the conscious use of cues.	Y	Use of cues are structured to enable graduation toward judgement application upon a decision. If cues provide a "supported" decision, then the user progresses no further. If cues are unable to be formulated into a sound argument, then the user progresses no further than Gate 2. If there is disgareement regarding the validity of the cues or debiasing is not possible, then the user does not progress beyond Gate 3.
	Presenting/displaying cue use	Y	The argument form requires the information/data/experiences/belief used to support their assertions. An influence diagram could support the argument by describing the cues that have bearing upon the decision choice, but this is not included in this version of the framework. The strength of the cues is assessed through when structuring the argument. Assumptions/uncertainties relevant to each cue are also explored through the argument rebuttals.
Justification	Mechanisms for justification	Y	The Toulmin argument structure provides the means to present justification in a consistent manner that is flexible to the available data. The heuristic gates provide explicit and specific questions that users must assert their position against. The gates provide triggers to identify the situation as supported/unsupported
	Individual judgement justification	Partly	Gates 2 and 3 address individual judgement use - firstly if a sound argument can be formed to support the judgement around the information/experiential knowledge available, and secondly if the judgement can be debiased or consensus around it formed. Individual perceptions are accounted for by virtue of the progression through the framework and the resulting argument formed around a judgement. **Framework does not require a threshold level of risk (quantified) to be established, only that a sound and strong argument that has consensus/is debiased.** Judgement/experience validity is assessed through the establishment of the argument.

System Requirements		*Development Model x-check*	
		v8	**Remarks**
Feelings and influences	Decision-maker values	Y	Values are inherently included by virtue of the cues that a decision-maker attends to and evidences in their argument. An initial non-deliberative "immediate perception" allows decision-makers to detail their reactive response (pertaining to Type 1 thinking) to the issue, in order to explicitly address biases that result from anchor values (anchoring, confirmation, availability, omission).
	Situational influence on intuition	Y	The framework is completed on the basis of the information available at the time. The heuristic tree used enable decision-makers to apply their judgement in a situationally dependent manner i.e. compelled to use unsupported judgement as a result of contextual factors.Gate 2 enables self-scrutiny on whether information used can form a sound argument, i.e. if operational pressures are bearing upon a decision, they may not provide a sound basis for an engineering risk argument. Gate 3 provides the opportunity for consensus for others to challenge the argument basis. **Spot or policy-type decision not explicitly accounted for, though the framework will tend to handle address a spot decision, but is dependent upon personal policies for their argument basis in application to a given situation.**
	Mechanisms to support	Partly	Gate 3 necessitates debias/consensus step before a judgement applied to a risky decision, providing confidence in judgement applied. **Toulmin argument structure provides mechanism for declaring the available information acted upon. Structuring judgement by arranging knowledge and information in an argument form will be tested as a means of supporting decision-makers.**
Model form	Provide Links: Info - cues - choice Risk - decision (values, agreement) Situation capture - d-m perspective - d-m honesty Decision - responsibility D-m frame - info available Info available - reflection on source Situational info - judgement Risk - judgement - true risk Model - supported RA - true risk	Partly	Framework does not provide an exhaustive description of the context, but instructs the user to define the relevant contextual factors that contribute to the argument. Heuristic tree narrows the frame of the situation (reliant upon user honesty of the situation), providing insight upon the information the user was prepared to act upon. Linking to the true risk is based upon perception only, which is why the inclusion of rebuttals within the argument structure is essential to indicate the uncertainties present within the mental model of the situation.
	Inherent features: Transparency Confidence Structures include information necessary for a decision Justification Handling supp/unsupp decisions Updatable with new info Barrier performance (APR) and interaction (PR)* Consistent language Systems perspective Heuristic - assessment perspective, env/context constraints, measurable criteria. Sociotechnical elements included Time perspective Mode of decision-making Argument	Partly	Transparency: documenting decision argument basis. **Confidence: will be tested** Justification: argument form provides the means. Supp/unsupp decision: the framework forms the basis of deciding to use supp or substantiating unsupp decisions. APR/PR: not directly addressed but may have x-over. Consistent language: does not use the SRA glossary as the framework focusses on decision architecture, rather than risk analysis, although terms will be aligned to glossary where appropriate in related material (training). Systems perspective: framework is constrained only by the user's perspective, so accounting for wider systems issues (human factors, policies, physical systems) is permitted by the framework. A lack of a handrail to guide users may lack wider appreciation of systems influences - education may support this. Heuristic features: criteria are individualistic and may not be objective - an experienced user may have the means to identify a supported solution (experience of contacts, location of information etc.), another may lack this ability and feel compelled to use judgement. The measurability of the criteria applied by a user is dependent upon their knowledge/skill. Sociotechnical system: allowing experiential influences to support argument accounts for sociotechnical factors. Time perspective: review number included on form to allow time bound assessments to be updated. **Mode of decision-making - not defined explicitly in the heuristic model, but accountable within the holistic process.** Argument: fundamental to the framework.
	Discrete features: Filter stage Show information available Show situational specifics Output is a decision Simplified structure/ease of recall Show assumption/uncert/cons Issue generalisability/uniqueness Encourage u/u, u/k,negligable event consideration* Compensatory factor identification Default to rule-based, unless overuled by RPD* Neutral framing Satisficed solution	Partly	Filter stage: Gate 1. Show information available: only information that supports the argument will be presented. **Show situational specifics: presented through the argument warrants and backing. Presently no account of operational pressures placed on the situation.** Output is a decision: heuristic framework outputs decisions. Simplified structure: three step process, "safe" argument metaphor to aid recall of argument structure. Show assumption/uncert/cons: accounted for through argument form (rebuttal). Issue uniqueness: related to spot/policy decision and wider applicability (fleet), which is part of the holistic process. **Encourage u/u/, u/k, negligable event consideration: not explicitly prompted for in this version.** Compensatory factor identification: heuristic tree fixes the significant factors, which are individualistically identified. Rule-based/RPD: framework is rule-based form, but relies upon experiential ability in recognising cues to form an unsupported argument. Neutral framing: "immediate perception" exposes the framing of the judgement. Satisficed solution: framework does not seek optimal quantified solution. Balances time to identify a decision, safety, and contextual state (information, time, resources).

Appendix J

Framework testing recruitment

Participant recruitment invitations

J.1 Practitioner survey (Phase 2) participant recruitment invitation

Invitation to participants for practitioner survey conducted in Phase 2.

"Dear structural integrity practitioner,

INVITATION TO STRUCTURAL INTEGRITY RISK PRACTITIONERS FOR PARTICIPATION IN RESEARCH SURVEY

You are invited to participate in a research survey that calls on your experience as a structural integrity risk practitioner and will provide valuable input into my research. My research is seeking to improve decision support for structural integrity risks in operational circumstances, and assist decision makers in justifying their judgements.

The invitation is purely voluntary and you are under no pressure or obligation to participate, but your contribution will be extremely valuable to the research. Any responses provided under pressure to complete the survey may harm the integrity of the research results.

A £20 Amazon voucher will be awarded in a random draw, for those survey participants who fully complete the survey and provide their preferred contact details. All personal details will be treated confidentially and strictly for the purposes of enquiries pertaining to this survey.

The link to the survey is as follows: *(link)*. The survey will close on *(date)*. The prize draw winner will be drawn by a member of Cranfield University staff independent of the study following closure of the survey, and will be

notified via the contact details provided through the survey.

Should you be aware of other colleagues or contacts that are practitioners of structural integrity, please feel free to forward on this invitation to them."

J.2 Framework testing (Phase 4) participant recruitment invitations

Invitation to participants for novel methodology testing during phase 4 who participated in an earlier research phase.

"Dear prospective participant,

Invitation to participate in aircraft structural integrity risk decision-making research - follow-up

Following a survey regarding aircraft structural integrity (SI) risk decision-making that you participated in between October and December 2019, you expressed an interest to participate in follow-up research. My research continues to aid the development of a novel framework to support decision-makers in situations where they are compelled to use their judgement. The survey that you contributed to was essential for identifying factors that were essential to consider in decision-making under operational constraints and identified findings such as:

a. Between 39-71% (95% confidence interval (CI)) of SI practitioners do not believe a risk likelihood is generated for SI risks.

b. Practitioners are split in opinion as to whether an acceptable threshold risk value is applicable to SI risks (between 39 and 72% (95% CI) participants believe they are required to assess SI risks relative to a threshold risk level).

c. All practitioners agree that human subjective judgement influences SI risk assessment.

d. Between 25-57% (95% CI) practitioners believe human subjective judgement is not communicated in risk assessments.

The research is now ready to proceed with the next phase of data collection.

You are invited to participate in the next data collection in order to further support the development of a novel framework for supporting

decision-making who are compelled to use their judgement. The next phase involves the assessment of a number of case studies, representing fictional but realistic operational scenarios in which a decision is required. The commitment required for the next phase of research will be between approximately 2 and 4 hours of your time, at a time of your choosing. The participation will be completed via web-based means and will be required to be completed before 27 Nov 20. Further details will be provided to you along with mandatory consent forms and participant information sheets following your returned agreement to participate in the next phase. Participation is wholly voluntary and you will be able to withdraw from the study at any time. Participants will receive a £10 voucher as a gesture of gratitude for their involvement.

The minimum requirements for the next phase of the research are that you must have experience of making decisions on aircraft maintenance structural faults under operational constraints.

If you would like to participate in the next phase, please reply to this email advising of your preferred contact means."

The invitation to non-Phase 2 respondents took the following form:

Invitation to new participants of novel methodology testing during phase 4.

"Dear prospective participant,

Invitation to participate in aircraft structural integrity risk decision-making research

Engineering managers are sometimes compelled to use their judgement when they lack the information, time or resources to objectively inform a maintenance decision. Recent research regarding judgement use in structural integrity risk decision-making has identified that this creates conflict with existing risk management systems. You are invited to participate in the research data collection, which will measure how practitioner maintenance decision-makers handle realistic but fictional decision scenarios. The participation is voluntary, will require between 2 and 4 hours of your time at a time of your choosing, and will be completed wholly by electronic/web-based means. Participants are able to withdraw at any time and participation is required to be completed before 27 Nov 20. Participants will receive a £10 voucher as a gesture of gratitude for their involvement.

Participant minimum requirements necessary are as follows:

- Trained engineering managers experienced in decision-making and

the management of structural integrity faults/issues.

– Experienced in decision-making on aircraft structural integrity engineering maintenance risks under operational constraints of time, resource and/or information.

Further details along with mandatory consent forms and participant information sheets will be provided to volunteer participants following agreement to participation.

If you would like to participate in the next phase, please reply to this email advising of your preferred contact means."

Appendix K

Framework testing scenario information

This appendix contains the information presented to Phase 4 participants. Section K.1 outlines the simulated scenario contextual and case study information presented to Sub-phase 4.1 participants. Section K.2 outlines the simulated scenario contextual information presented to Sub-phase 4.2 participants. Section K.3 details the questions asked of both sub-phases during the novel framework testing. The results were described in Chapter 6.

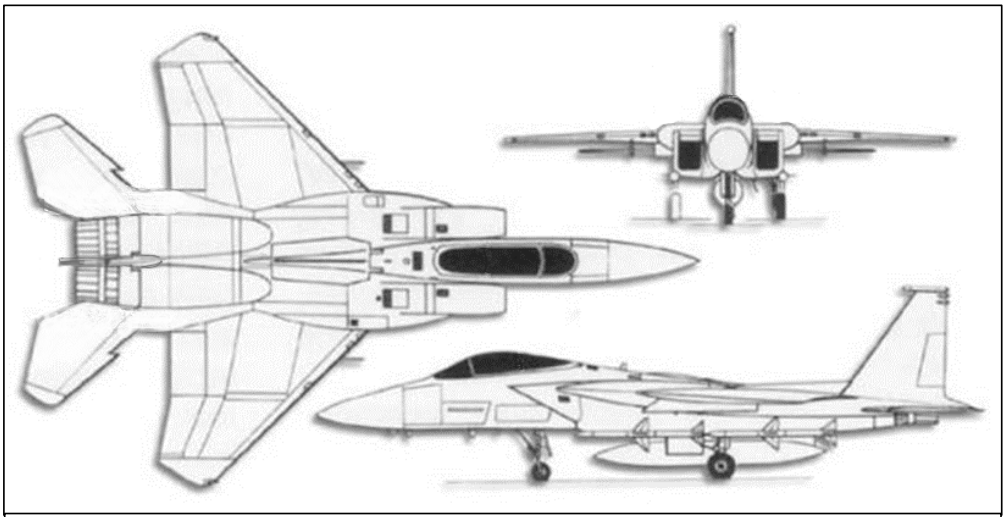
K.1 Phase 4.1 participant context and case study information

CONTEXT

The subject aircraft and scenarios are fictional but realistic. The 4th generation Rapier Storm is a single seat fighter aircraft that operates a pressurised cockpit and has a twin-engine configuration. There are 100 aircraft in the fleet, 10 of which are used for operational conversion training and tend to show higher fatigue usage and landing totals but lower flying hours compared to the remainder of the fleet. Design features relevant to each case study will be detailed. The aircraft was initially designed to perform for 5000 flying hours using Safe Life principles, with some critical components being assessed using a damage tolerant design philosophy. The fleet has since been extended to 8000 flying hours.

In all cases, you are an engineering manager responsible for aircraft structural integrity of the fleet. You are required to maintain the airworthiness[1] of the fleet, providing balanced risk assessment of engineering faults against the operational requirement to fly, you are delegated with appropriate authorisations to make these decisions. You operate within your current regulatory and policy constraints. Your line management look to you to decide upon whether aircraft (individual or fleet wide) should be grounded, or limited in performance, in response to alerts about aircraft faults.

For the purposes of the case studies assume that, if necessary, you have a similarly experienced colleague (Mrs A. Bloggs) as yourself who is in agreement about your assessments.



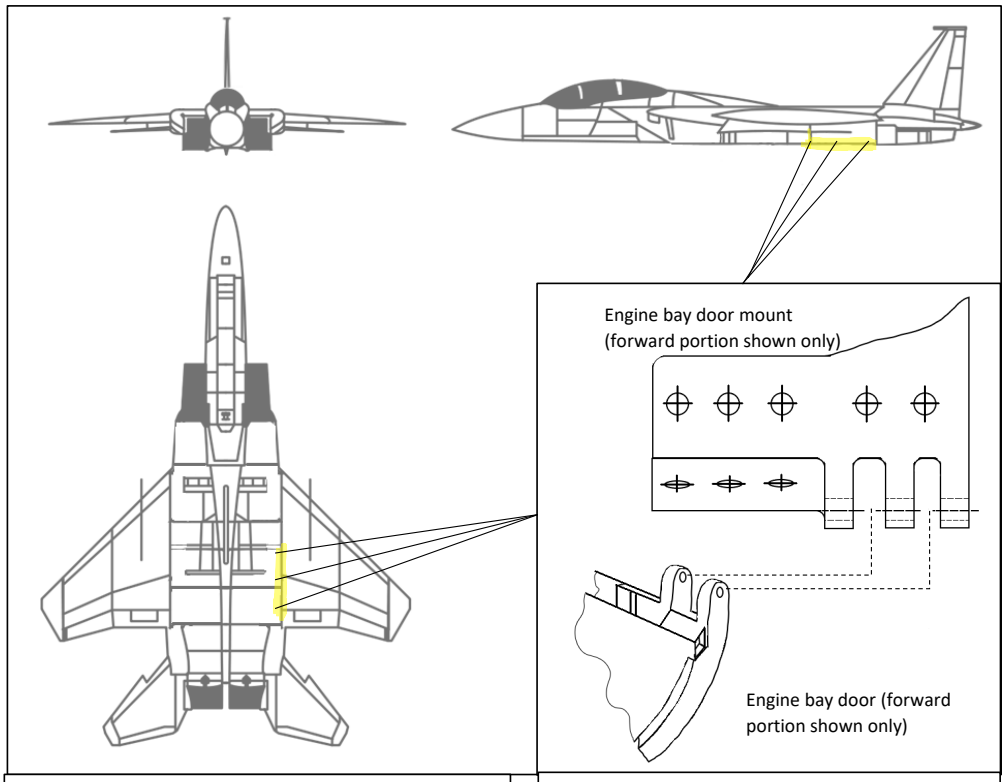
Context, Figure 1 – Aircraft general arrangement. *This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original.*

[1] "Airworthiness" is a measure of an aircraft's suitability for safe flight.

CASE 1 - ENGINE BAY DOOR ATTACHMENT LUG WEARING.

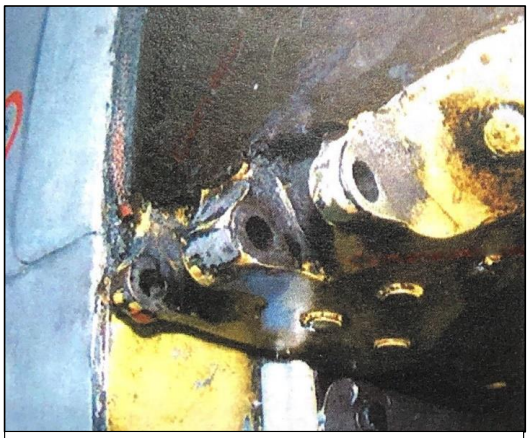
You receive an emailed request for airworthiness advice from the maintenance unit responsible for schedule maintenance. During a depth maintenance activity, an aircraft in the Rapier Storm fleet has been identified as having out-of-limits wear to the attachment lugs which the engine bay doors are mounted to. All attachment lugs across the three engine bay door mounts have wear in the lug holes, signs of wear and not permitted according to the aircraft documentation. Maintenance documentation does not appear to list the mounting points as primary structure, but the engine bay doors are primary structure. The mounting point is safe life limited, and is cleared within the wider aircraft clearance to 8000 flying hours. A telephone call with an OEM design engineer responsible (a trusted and experienced for the rear fuselage) confirms that the material has a short fatigue critical crack length of "a few millimetres" but there are no visible cracks at present. NDT services may be able to confirm the presence of non-visible cracking but the engine bay doors are required to be removed (12 man hours labour) to have sufficient access for a sufficient NDT examination of the lug holes. A second aircraft undergoing maintenance has also been found to have similar wear. Both aircraft are mid-life and not the fleet leaders in usage (flying hours, landings, or fatigue usage). No previous occurrences of this fault have been recorded within the Rapier Storm fleet and experienced colleagues cannot recall seeing the fault previously. An urgent request for OEM repair instructions will take at least 7 days to produce and distribute, and replacement of the damaged lugs is not possible without replacement of the main rear fuselage loading bearing frame, spares for which have an extremely long lead time and would be an intensive maintenance task. There are 6 aircraft currently deployed on operations support overseas (NDT available) that are flying daily, and 4 taking part in international exercises (NDT not immediately available). The remainder of the fleet are in the UK either in maintenance or allocated to front line squadrons for routine training sorties. You are required to make a decision in the next 20 minutes on how to handle this risk with regards to the remainder of the fleet.

Scroll down for figures…



Case 1, Figure 1 – General arrangement showing damage location. *This Photo by Unknown Author is licensed under CC BY-SA, and has been modified from the original by Richard Green.*

Case 1, Figure 2 – Engine bay door mount configuration (forward), x3 total. Similar configuration for others



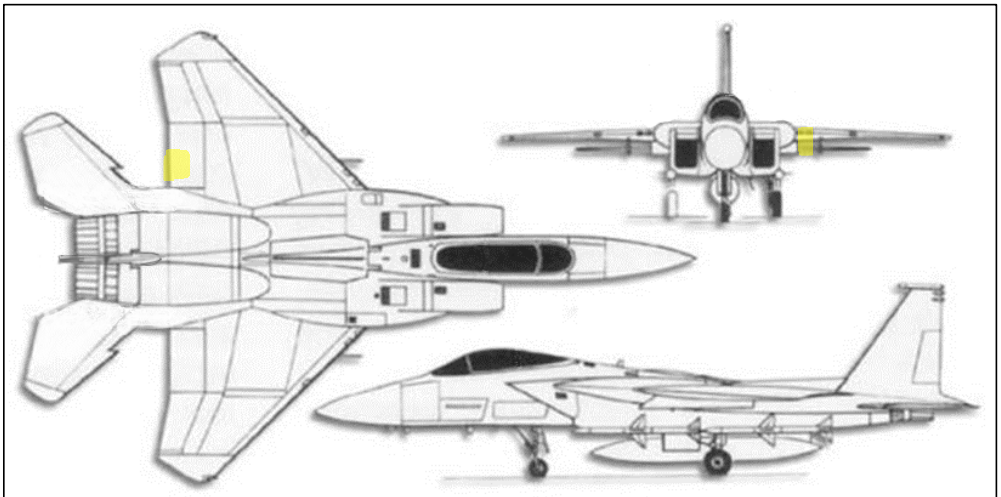
Case 1, Figure 3 – Forward port side engine bay door mount.



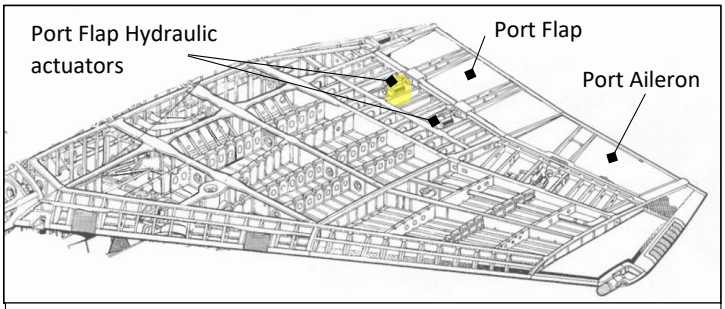
Case 1, Figure 4 – Centre port side engine bay door mount.

CASE 2 - FLAP MOTORING MECHANISM ATTACHMENT BOLT.

During a scheduled maintenance to an aircraft on an operational conversion unit, while replacing a flap a technician found a sheared retaining bolt that attaches the port-side inboard flap actuator to the flap. The bolt was failed through thickness, to the extent that the bolt would no longer be securely attached to the flap but the flap actuation mechanism was found still attached to the flap, despite the bolt failure. On physical inspection of the failed bolt, there is surface corrosion present, identifiable by discoloration on the fracture surface. There are no indications of consequential damage to the flap or mating wing structure, such as a fretting or wear, and there have been no recent airborne indications of flap failures attributable to retaining bolt failures. There are 2 actuation attachment points per flap, and one flap on each wing, meaning there are 4 similar bolts across the flap system on the aircraft. The technical documentation does not provide any allowable limits for failed retaining bolts or flap motoring mechanism. The aircraft is used for operational conversion training and is one of the fleet leaders for total landings, but is within the designed limit for total landings, and is midlife in other life monitoring metrics (flying hours, pressurisations, full stop landings). No other failures of this type have been identified before and there is no immediately apparent history on the aircraft of reported flap control issues, asymmetry or fault codes. Formal OEM advice will take at least 7 days. In discussion with an experienced and trusted technician, you assess that to visually inspect a single retaining bolt will take approximately 0.5 man hours of work. Replacement bolts are available but there are insufficient quantities to conduct a 100% replacement program for the aircraft in the flying fleet – there are 40 replacement bolts currently available, and new orders taking 10 days to fulfil. There are 10 OCU aircraft, and 90 forward fleet aircraft, 6 of the latter are deployed on operations supporting active military operations daily. The OCU is required to continue operating and a period of grounding will further delay the throughput of trainee aircrew, which is already behind schedule. You are expected to make an immediate decision on how to handle this risk with regards to the remainder of the fleet.

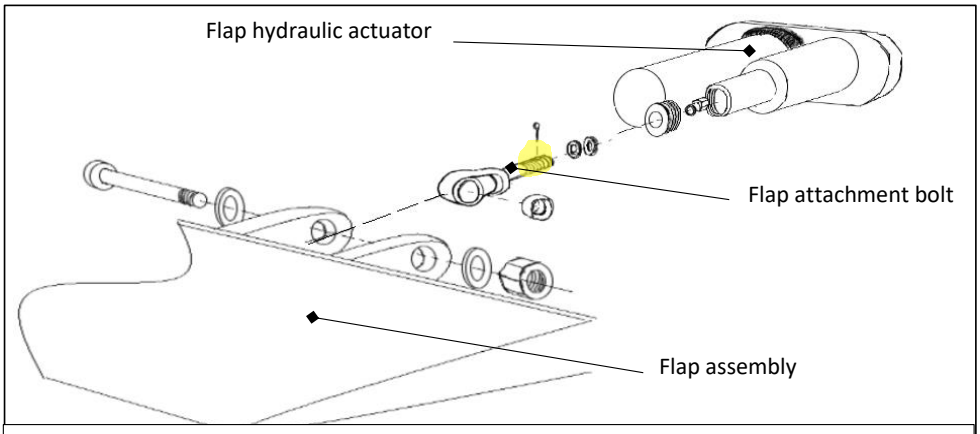


Case 2, Figure 1 – Aircraft general arrangement showing damage location. *This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original.*



Case 2, Figure 2 - Port side wing view* indicating location of flap motoring mechanism.

* This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original by Richard Green.



Case 2, Figure 3 - Exploded view of flap motoring mechanism assembly, indicating the cracked area of the threaded portion of flap attachment bolt



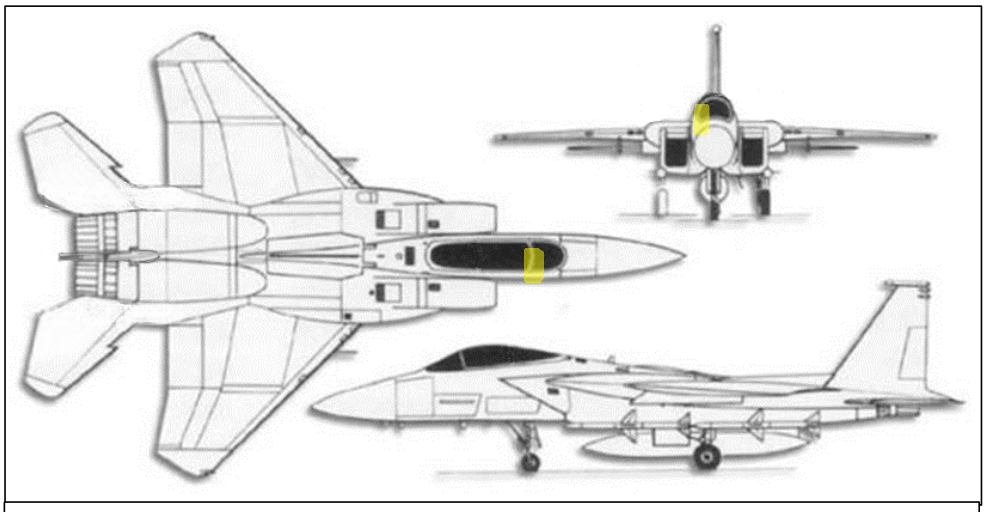
Case 2, Figure 4 - Flap attachment bolt failure, indicating failure initiation from the split pin hole.



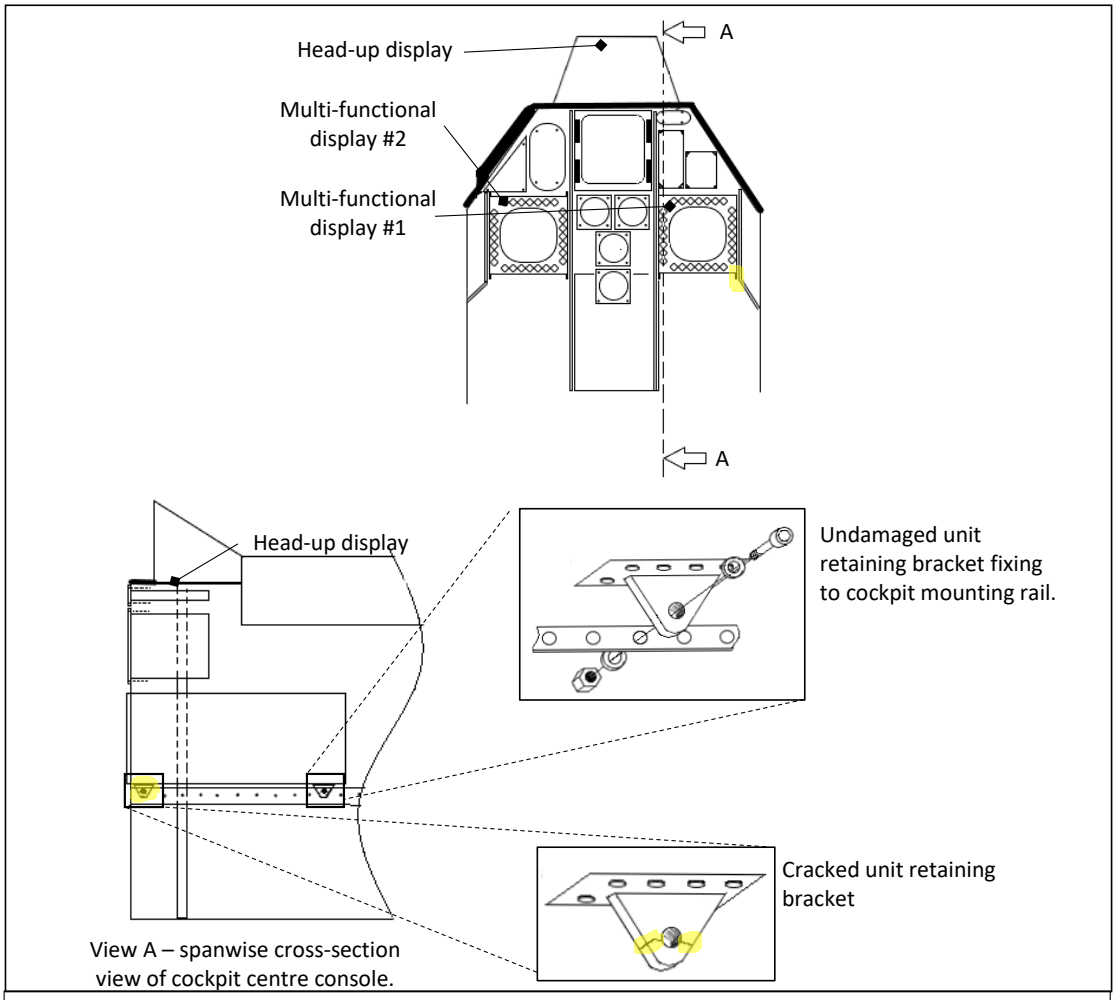
Case 2, Figure 5 - Flap attachment bolt failure, showing corrosion discolouration to fracture surface.

CASE 3 - COCKPIT INSTRUMENT RETAINING BRACKET.

Following aircraft shutdown on a forward operating squadron, aircrew reported to maintainers that an instrument unit had come loose in their aircraft's cockpit during flight but had not fully detached. The issue has been brought to your attention after it was reported by the aircrew through the air safety reporting tool. The instrument is mounted on the central instrument console above the pilot's knees and was reported to be "wobbling" during flight. Inspection by technicians following removal of the instrument identified one of the four aluminium mounting brackets that attaches the corners of the instrument console to be fractured through thickness, with the fracture line running radially through the thickness of the bracket. There were no visible signs of damage to the 3 remaining brackets, confirmed by NDT. The broken bracket was replaced by technicians and the aircraft is continuing to operate without limitation. Visible inspection requires removal of the instrument panel and an experienced colleague estimates 3 maintenance hours to inspect all four brackets and recover each aircraft. No other faults of this nature have been identified previously and the aircraft is a young aircraft in the fleet across all major metrics – flying hours, landings, pressurisations. The mounting bracket has not been designated with a specific life measuring unit, meaning its life is the overall aircraft life limit, which is not believed to have been exceeded as the brackets are not transferred between aircraft. The instrument has not been changed any more regularly than on other aircraft in the fleet. Formal OEM advice is anticipated to take at least 3 days. The aircraft is part of the forward flying fleet conducting routine UK training, but the operators are pressing to maintain as many aircraft flying as possible. You are required to make a decision in the next 20 minutes on how to handle this risk with regards to the remainder of the fleet.



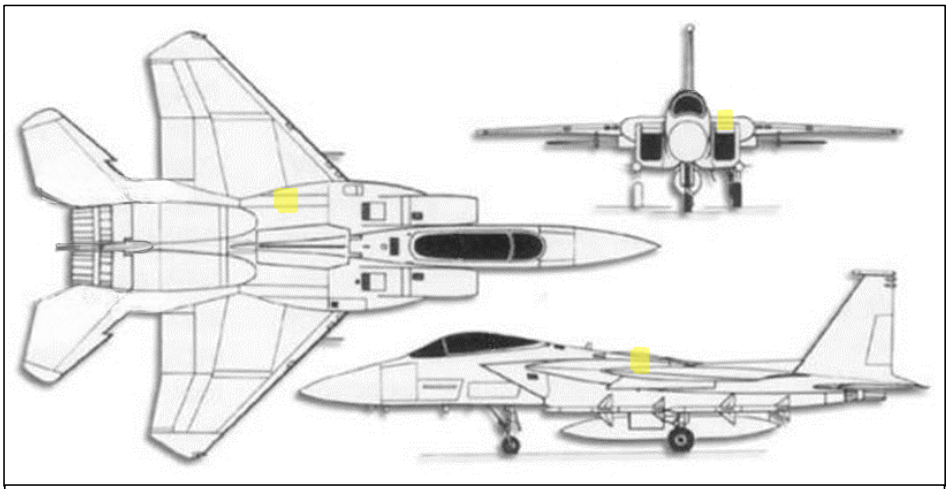
Case 3, Figure 1 – Aircraft general arrangement showing damage location. *This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original.*



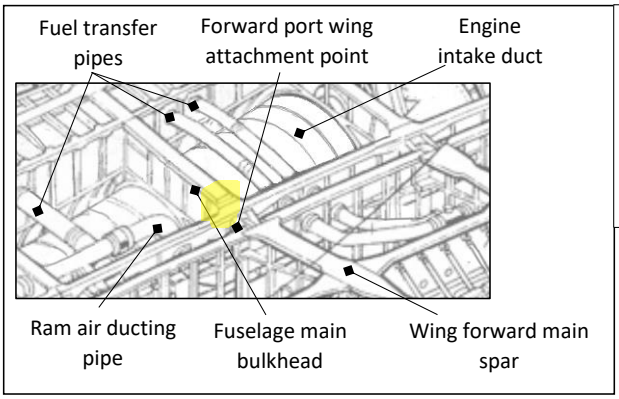
Case 3, Figure 2 -Cockpit front centre console arrangement looking forward, showing view A and indicating failed bracket location.

CASE 4 - TOOLING DAMAGE SUSTAINED ON PRIMARY FRAME STRUCTURE.

You receive urgent correspondence from a deployed maintenance manager. During a front-line maintenance task on a deployed aircraft to replace a fuel system component within an accessible bay on the upper port-side of the centre fuselage, tooling damage in the form of two small dents has been imparted onto a major load-bearing titanium fuselage frame and a mating lap joint. The frame/joint arrangement is one of three that transfers load between the port wing and the fuselage through the three wing attachment points. The damaged frame is the forward main fuselage bulkhead (of three). The frame and joint are primary structure and the dents are seemingly out of limits in accordance with maintenance manuals, as no limits are provided for this frame. Discussion with experienced tradesmen reveals that standard repair techniques have previously been applied to similar damage sustained on other frames on the aircraft, mechanically blending the damage out and conducting NDT on the repaired area to confirm no presence of cracking following the blending. However, your team do not recall damage to this frame previously and NDT services will be on-site in 48 hours. The damage is easily accessed by removal of a skin panel. The frame and lap joint are not replaceable at the front-line as the components transfer primary wing loads and would require wing removal, a task for which front-line maintenance is not capable. Replacement of the lap joint (but not the frame) would be possible during depth maintenance, the next of which is in 150 flying hours. The Squadron have consequentially reduced their aircraft availability for operations, but the operational commanders have communicated that this is influencing your organisation's reputation in the theatre of operations. You are required to make a decision in the next 20 minutes as to how to handle this risk with regards this risk.

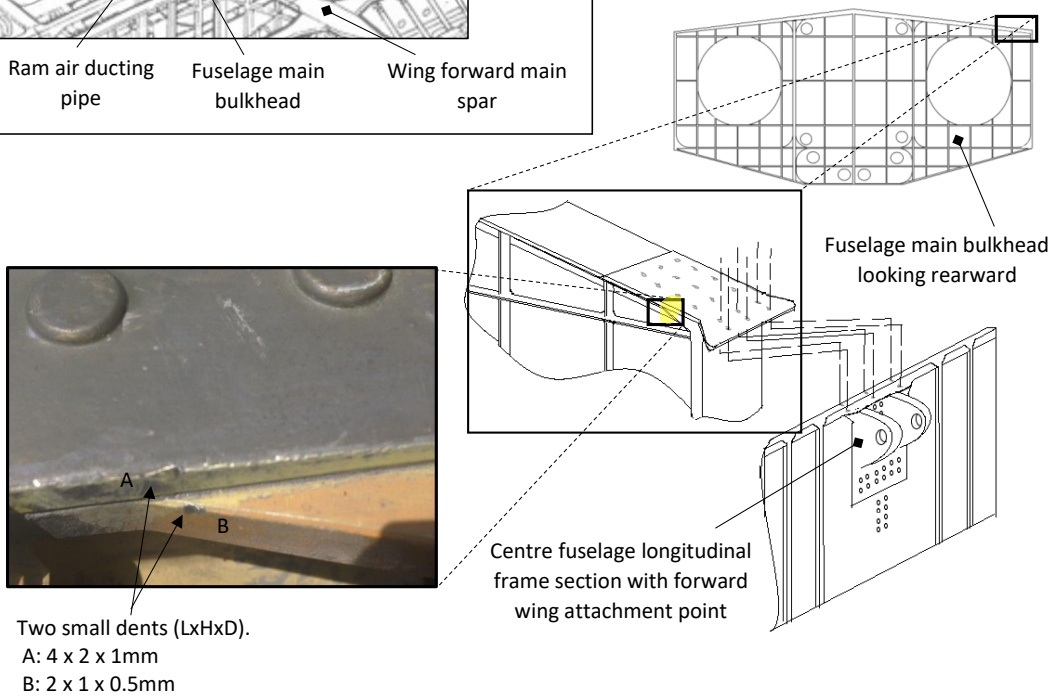


Case 4, Figure 1 – Aircraft general arrangement showing damage location. *This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original.*



Case 4, Figure 2 – View of wing/fuselage joint and intake duct zones exposed with skin panels removed.

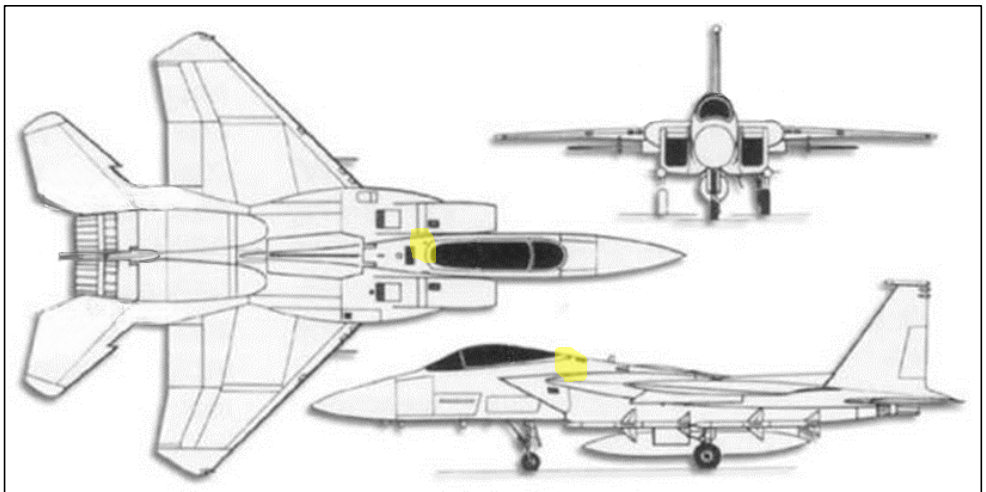
* *This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original by Richard Green.*



Case 4, Figure 3 – Photograph of two small dents to bulkhead frame, and indication of damage location on bulkhead frame.

CASE 5 - COCKPIT CANOPY HINGE BUSH.

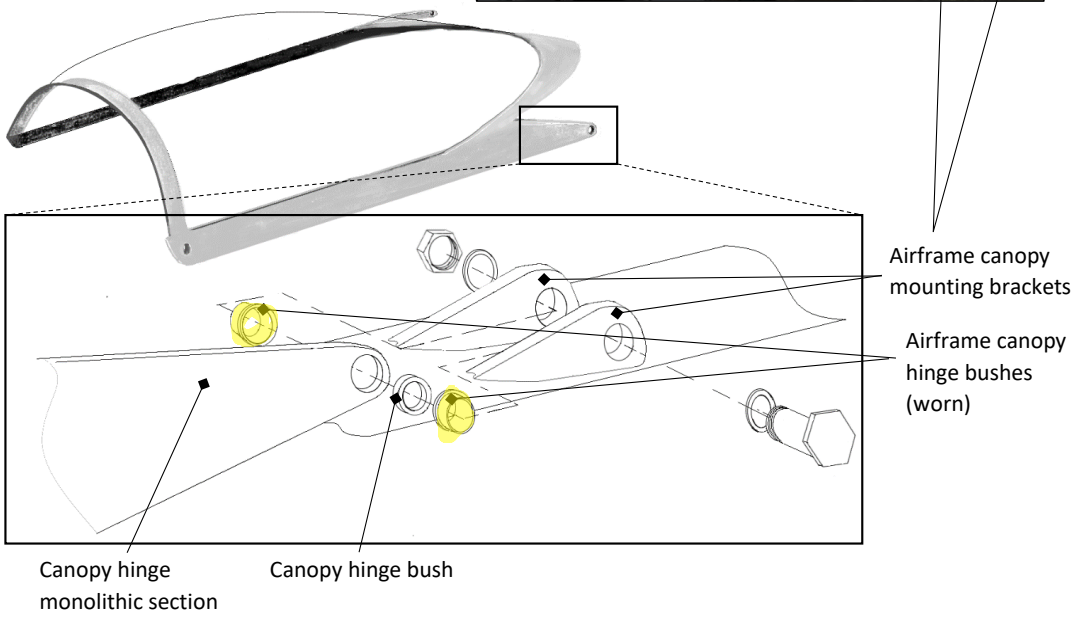
You receive a routine request for repair advice from a front-line squadron at a main operating base. During an aircraft despatch, the canopy would not give “fully closed” indications in the closed position. Maintenance investigation found the canopy hinge bushes excessively worn beyond approved wear limits (published in technical documentation), but the technicians were also able to functionally operate the canopy satisfactorily in accordance with functional testing procedures. The aircraft is not within the 10 fleet leading aircraft in any structural metrics and there is no published life for the bush, meaning it adopts the life of the aircraft by default. The squadron have stated in their report that they are continuing operating the aircraft owing to “operational necessity”. A previous repair instruction is available and approved by the OEM for this type of fault which is believed to occur as a consequence of taxiing and towing the aircraft with the canopy open without ground locks in-place, a discouraged practice which continues to be practiced by both aircrew and groundcrew. Replacement bushes must be manufactured on-demand and will not be available for another 48 hours (either locally manufactured or outsourced). NDT services are available on demand at the aircraft location and the maintenance task to remove the canopy to gain access to the bushes is estimated to take 4 man-hours. Replacement of the canopy hinge bushes is estimated to take 6 man-hours but your experienced colleague advises that the old hinge bushes are notoriously difficult to remove. The demand for the aircraft is high as the Squadron are in deployment ‘work-up’ with a much higher flying-rate to ensure the aircrew are prepared adequately for operational flying and to handle enemy threat. You are required to make a decision in the next 20 minutes as to how to handle this risk and scrutinise the decision by the Squadron to continue operating the aircraft.



Case 5, Figure 1 – Aircraft general arrangement with damage location indicated. *This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original by Richard Green.*

Case 5, Figure 2 – Photo of port side airframe canopy mounting brackets and surrounding structure.

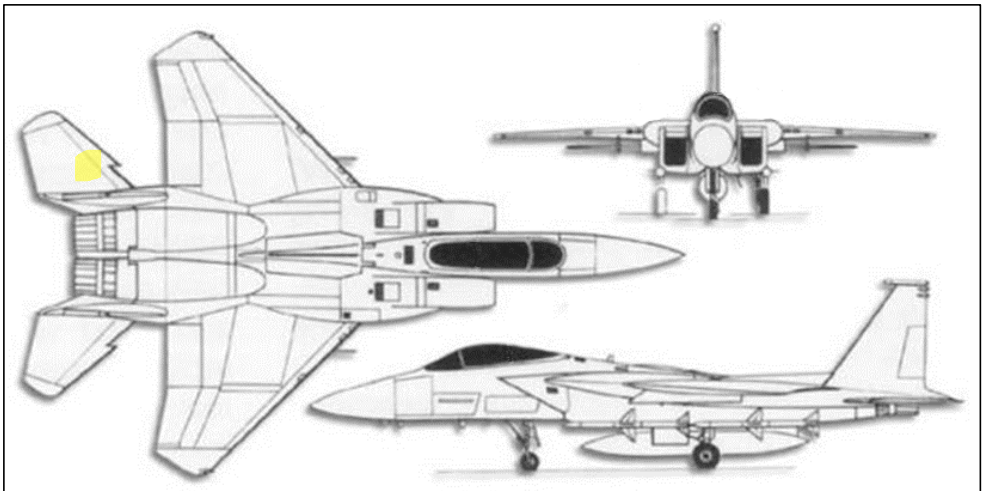




Case 5, Figure 3 - Exploded arrangement view of canopy hinge assembly attachment to fuselage

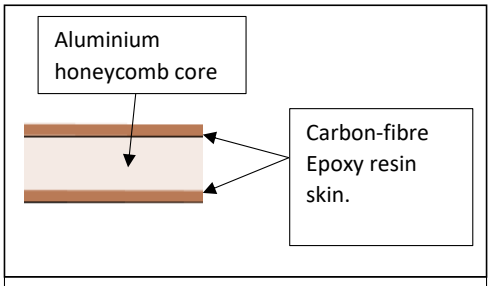
CASE 6 - NUMEROUS DENTS ON ELEVATOR UPPER SURFACE.

5 small dents on the upper surface of the port tailplane have been reported to you by a front-line squadron following a post-flight servicing during an international exercise taking place overseas. The dents are all within the same centre portion of a skin panel away from rivet lines, which has an epoxy-carbon composite sandwich construction. The squadron are uncertain how long the dents have been in-situ, as they were only spotted because an unrelated post-flight check of another component required access to the upper surface of the tailplane. Taken individually, the dents are all within damage limits set by the maintenance manual, but the cumulative number and proximity to one another means that the damage is considered out of the specific damage limits published for dents on the tailplane. The squadron have taken a pencil rubbing of the dents to provide dimensions, and have carried out visual inspections as far as possible observing that there is no other damage apparent. NDT services have been requested and will be on location within 24 hours to support if required. The starboard tailplane is undamaged. Aircrew have not reported handling or control concerns with this aircraft. Your experienced team do not recall any identical damage on other aircraft, although dents have been handled on sandwich construction control surfaces previously by applying resin-fill and bonded repair patch repairs in accordance with the structural repair manual, or by replacement. A spare tailplane is not immediately available at the aircraft location but can shipped and received by the squadron in 7 days. Replacement of the damaged panel on the tailplane is not possible with the personnel and tooling available on location owing to the complexity of the task. The tailplane was replaced at the last major servicing 312 flying hours/20 months ago. Seeking formal OEM advice regarding whether it is safe to continue flying within the exercise until the aircraft returns back to home base is expected to take between 3 days. The aircraft is due to return to main operating base in 20 days. The squadron have communicated that they intend to continue operating the aircraft to maximise the training opportunity presented by the international exercise but have notified you of the damage as it is outside of approved limits for damage to the tailplane. You are required to make a decision within the next 20 minutes on whether to endorse the squadron maintenance team's actions to continue operating.

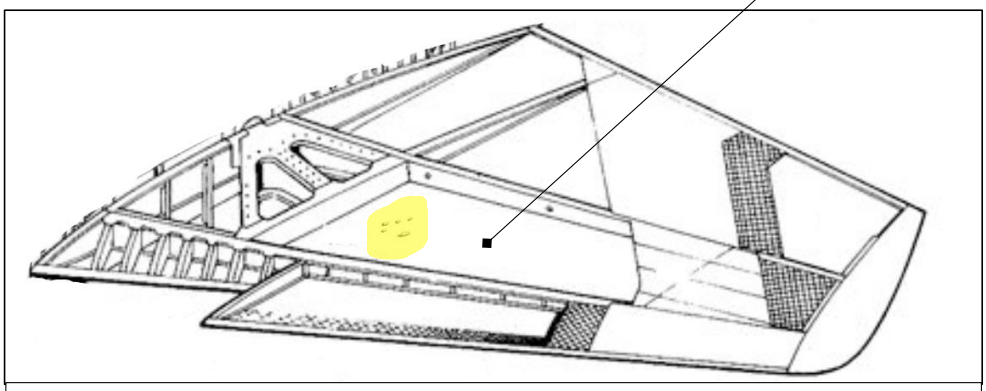


Case 6, Figure 1 – Aircraft general arrangement with damage location indicated.

This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original by Richard Green.

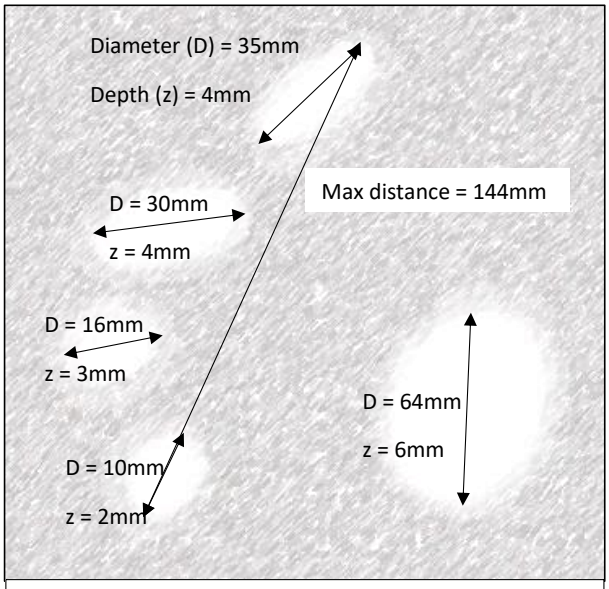


Case 6, Figure 3 – Tailplane skin sandwich panel construction at damage location.



Case 6, Figure 2 – Tailplane cutaway diagram with damage location indicated.

This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original by Richard Green.



Case 6, Figure 4 – Tailplane dent dimension details provided by pencil “rubbing”.

K.2 Phase 4.2 participant context information

PHASE 4.2 TASK OUTLINE AND CONTEXT

You will be asked to compare 6 independent pairs of decision descriptions/justifications produced by real decision-makers in an earlier phase of the research for 6 fictional but realistic operational military aviation maintenance scenarios. Each case study corresponds to a fictional military aircraft fleet, the overarching context for which is given below; you will be given a brief scenario-specific context prior to each decision justification pair. There is no relationship between the scenarios, and each should be considered as independent. You will be asked to select which of each decision description/justification pair is easiest to understand, which is most transparent, and which offers a better justification of the decision given. You are not required to consider which you think is correct.

The following definitions are provided to assist you in understanding the decision descriptions; you may wish to refer to this file throughout your participation:

Definitions:

Gate 1: requires the decision-maker to consider whether it is necessary for them to use their judgement in the situation that they are in (see figure 1a).

Gate 2: requires the decision-maker to consider whether they are able to form a sound argument from the information and knowledge they have available (see figure 1a).

Gate 3: requires the decision-maker to consider if there is consensus for their decision justification to provide a debiasing function (see figure 1a).

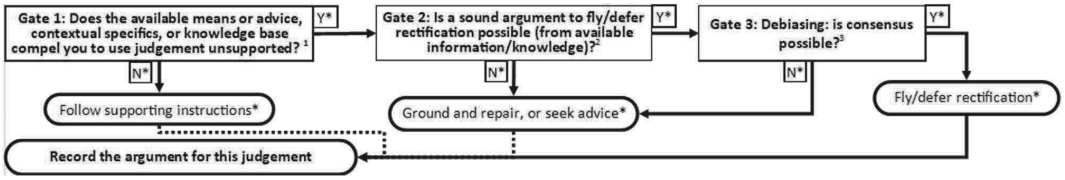


Figure 1a – Decision gates and their terminations.

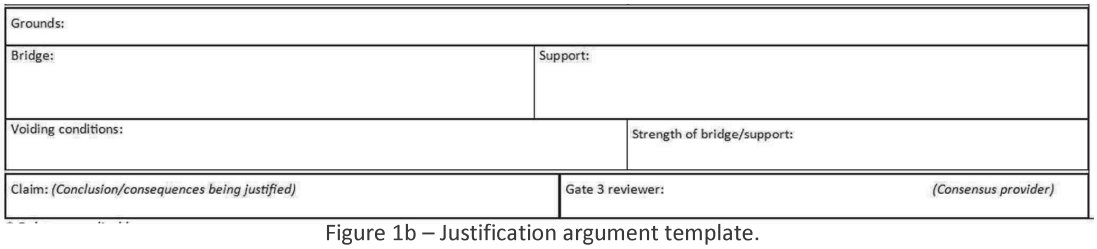


Figure 1b – Justification argument template.

Grounds: details the evidence, information or data used to initiate a statement in support of a *Claim* (see figure 1b).

Claim: the final conclusion drawn about the situation (see figure 1b).

Bridge: the "step" or linking clause between the *Grounds* and *Claim* (see figure 1b). There may be multiple sub-clauses within the bridge that enable the *Grounds* to conditionally be linked to the *Claim*.

Support: generalised or categorical information that provides support to the *Bridge* if deemed necessary (see figure 1b). There may be multiple sub-clauses within the *Support* to enable the *Bridge* to be sufficiently supported.

Voiding conditions: assumptions, uncertainties or conditions that, if they occur, would void the truth of the *Claim* (see figure 1b).

Strength of support: indicates a measure of the strength of the argument (see figure 1b). Guide words are given below, but you may encounter individualistic (i.e. personal to the decision-maker) variations.

> "*Strong*": a link between the clauses (*Grounds*, *Bridge* and *Support*, *Claim*) is known to exist and is applicable to the situation being considered. There is a relevant causal chain or correlation that is accepted by consensus or authority.
>
> "*Some*": the link is not strong, i.e. a link between the clauses is possible, but correlation or causality is not accepted by consensus or authority, or may be irrelevant to this situation.
>
> "*No*": there is no clear link between the clauses able to be formed from available information, knowledge or data.

Glossary and abbreviations:

The following terms may be seen in the decision descriptions/justifications that you're given:

Glossary and abbreviations	
71IR Sqn	71 (Inspection and Repair) Squadron – a bespoke structural repair and inspection unit within the Royal Air Force, who deploy specialist personnel worldwide to carry-out complex structural repairs or non-destructive testing
Ac	Abbreviation of "aircraft"
ADF	Acceptable Deferred Fault – deferment of maintenance rectification
ADS	Aircraft Document Set
ALARP	As Low As Reasonably Practicable – legally meaningful term indicating that the level of risk of a given situation is reasonable given the costs (time, money or trouble) associated with further reduction of the risk.
BVD/BVID	Barely Visible Impact Damage – composite damage mechanism.
CAE (DDH)	Chief Air Engineer to the Delivery Duty Holder - named accountable individual for advising a duty holder on engineering matters.
CAMO	Continuing Airworthiness Monitoring Organisation
CTI	CAMO Technical Instruction – technical instruction released by the aircraft CAMO.
Ctl	(as in secondary fg ctl surface) – control.
DAOS	Design Approved Organisation System – regulatory approved design organisation.
DASOR	Defence Air Safety Occurrence Report – reporting system for all air safety related occurrences within the UK MOD aviation domain.
DDH/ADH	Delivery Duty Holder - named accountable individual for the safety of a particular fleet of aircraft.
EA	Engineering Authority - military organisation responsible for fleet airworthiness and supportability through-life.
EngO	Engineer Officer – engineering maintenance manager, usually authorised delegated airworthiness decision-maker.
F760	Defect Fault Report form – military reporting form requesting detailed investigation into a specific fault occurrence on a component.
FFP	Fit for Purpose
Fg	Abbreviation of "flying".
FJ	Fast Jet

Glossary and abbreviations	
LA	Loose Article - items unaccounted for that a possible in the aircraft which are hazardous because they may cause issue in flight such as through restricting control runs.
LoAA	Letter of Airworthiness Authority – an official letter conferring delegated airworthiness decision-making responsibility to a named individual on specifically named systems of an aircraft type.
MFD	Multi-function display – cockpit instrument that can be selected to display different information screens.
MilCAM	Military Continuing Airworthiness Manager - responsible for continuing airworthiness matters
MOB	Main Operating Base
NDI/T	Non-Destructive Inspection/Testing
OC Eng	Officer Commanding Engineering and Logistics Wing - military engineering senior manager role.
OCU	Operation Conversion Unit - training for aircrew to convert on to aircraft type.
OEM/DA/DO	Original Equipment Manufacturer / Design Authority / Design Organisation
OEU	Operational Evaluation Unit – a unit that flight and ground tests aircraft and equipment, for example when introducing modifications to an existing aircraft.
Op	Operation – flying mission, may refer to a single flight or a military campaign.
OSD	Out of Service Date - expected termination date of the flying fleet.
PSG/PSP/ ASSWG	Various aircraft fleet safety working groups.
RTB	Return to Base
RTI	Routine Technical Instruction – technical instruction with compliance interval measured in weeks.
SI(T)	Special Instructions (Technical) – technical information released as an interim to full publication within technical manuals.
SOPs	Standard Operating Procedures
Sqn	Squadron
TAA	Type Airworthiness Authority - head of the EA for a fleet including financial and administrative aspects.
TI	Technical Instruction – encompassing all authorised technical documentation released outwith the primary technical manuals.
U/S	Unserviceable
UTI	Urgent Technical Instruction – technical instruction compliance interval measured in days (released with greater urgency than RTI).

Overarching Context:

The subject aircraft and scenarios are fictional but realistic. The subject aircraft is a 4th generation single seat fighter aircraft, Rapier Storm, that operates a pressurised cockpit and has a twin-engine configuration. There are 100 aircraft in the fleet, 10 of which are used for operational conversion training and tend to show higher fatigue usage and landing totals but lower flying hours compared to the remainder of the fleet. The aircraft was initially designed to perform for 5000 flying hours using Safe Life principles, with some critical components being assessed using a damage tolerant design philosophy. The fleet has since been extended to 8000 flying hours.

In all cases, the decision justifications provided are made by a non-deployed engineering manager responsible for aircraft structural integrity of the fleet. A brief introduction to the issue in each case is provided, but the details pertaining to the situation and the decision will be contained in the individual decision justifications.

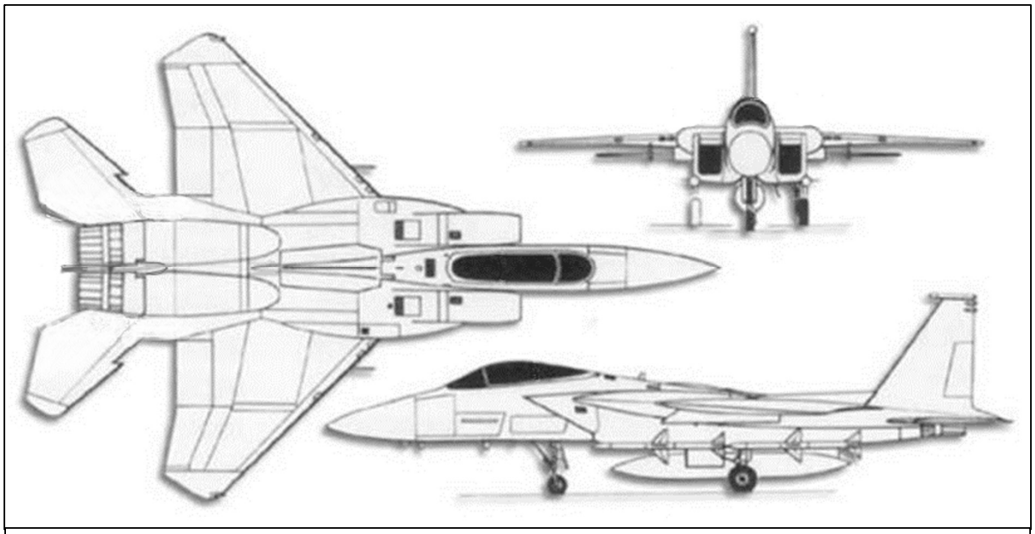


Figure 2 – Aircraft general arrangement. *This Photo by Unknown Author is licensed under CC BY-SA, and has been altered from the original.*

K.3 Framework testing participant questions

K.3.1 Question set for participants in Sub-phase 4.1

Context questions

1. How long have you been a practicing engineering manager?
   [$<$1 years / 1-5 years / 5-10 years / $>$10 years]

2. When was the last time you used your judgement to support an engineering risk decision when under operational constraints of time, resource, or information in the course of your employment?
   [In the last 7 days / In the last month / In the last 6 months / In the last year / In the last 3 years / Not within the last 3 years / Never]

3. Have you previously received training in how to make judgement-based engineering risk decisions in the course of your employment?
   [Yes / No]

Questions after each case study

1. Regarding the last case study answered only, how confident did you feel in presenting your judgement on the risky issue?
   [Extremely confident / Very confident / Moderately confident / Slightly confident / Not confident at all]

2. Regarding the last case study answered only, were there any aspects of your judgement that you were unable to communicate?
   [Yes / No]

   Please specify.

Reflection questions

1. How far do you agree with this statement: this method I applied for justifying my judgement would be usable in an operational engineering environment to support judgement-based risk decision making?
   [Strongly agree / Agree / Somewhat agree / Neither agree nor disagree / Somewhat disagree / Disagree / Strongly disagree]

   Why did you give this answer?

2. What, if anything, did you find unhelpful about the method presented?
   [Open answer format]

3. What, if anything, did you find helpful about the framework presented?
   [Open answer format]

K.3.2 Question set for participants in Sub-phase 4.2

Context questions

1. How long have you been a practicing engineering manager?
   [<1 years / 1-5 years / 5-10 years / >10 years]

2. When was the last time you used your judgement to support an engineering risk decision when under operational constraints of time, resource, or information in the course of your employment?
   [In the last 7 days / In the last month / In the last 6 months / In the last year / In the last 3 years / Not within the last 3 years / Never]

3. Have you previously received training in how to make judgement-based engineering risk decisions in the course of your employment?
   [Yes / No]

Scenario decision justification comparison Case study answers from Sub-phase 4.1 were presented for participants to select their preference according to the questions below. The case study answers presented were codified with a generic label (Answer X and Answer Y) to reduce the opportunity for respondent bias.

1. Which, of the two decision justifications presented, is the easier to understand?

   a. To what extent is the option you chose easier?
      [To a very little extent / To some extent / To a great extent / To a very great extent]

2. Which, of the two decision justifications presented, provides the greater transparency of the decision basis provided by the author?

   a. To what extent does the option you chose provide greater transparency?
      [To a very little extent / To some extent / To a great extent / To a very great extent]

3. Which, of the two decision justifications presented, offers a better justification?

   a. To what extent does the option you chose offer a better justification?
      [To a very little extent / To some extent / To a great extent / To a very great extent]

Appendix L

Novel framework training manuals

L.1 Full Training Manual

AIRCRAFT ENGINEERING MAINTENANCE

JUDGEMENT-BASED DECISION-MAKING SUPPORT FRAMEWORK

FULL TRAINING PACKAGE V1.1

TRAINING PACKAGE

Context

Aircraft engineering managers oversee multiple systems and it is common for numerous time critical decisions to compete for a manager's time. Some situations compel engineering decision-makers to use their judgement, where waiting for supportive resources or an analytical response from authoritative sources would be unacceptably time consuming. Decision-makers must choose whether to declare aircraft as safe or unsafe. Risk management systems necessitate the appropriate handling of hazards and risky prospects and enable the use of quantitative thresholds to instruction decision-making interventions. But the complexity of structural integrity issues means that the generation of a meaningful quantified risk for a newly apparent fault is perceived as an impossibility by decision-makers. However, there is presently no guidance on how best to apply experience and judgement to uncertain and ambiguous structural faults in order to decide on how best to proceed. This framework attempts to bridge into this unknown.

There are two distinct components to the framework: a *heuristic tree* that rapidly terminates to a decision; and an *argument structure* that presents the user's argument regarding the situation. Using these two methods supports judgement-use in maintenance decision-making because they enable the bits of information used in a judgement-based argument to be laid out transparently. This method of risk assessment provides an intuitive risk communication tool until a relevant and meaningful supporting assessment is available, such as from OEM analysis. Instruction on the use of these is provided in this document.

Rich Green
rich.n.green@cranfield.ac.uk

THEORY

Argument

Considering aviation safety specifically, regulatory and accountability requirements necessitate transparency in decision-making; "argument" provides a vehicle to do so for judgement-based decisions. Argument formulisation provides an unambiguous structure around which users can arrange the information that they are basing their claim about. For example:

Premise 1: All birds have wings

Premise 2: A penguin has wings

Claim: So, a penguin is a bird.

A claim without basis is weak and would not stand scrutiny, so removing one of the premises above would not allow the conclusion "a penguin is a bird" to withstand logical scrutiny. But arguments can also be challenged by the advent of new information, for instance:

Premise 1: Most birds can fly

Premise 2: All penguins are classified as birds

Claim: So, a penguin can probably fly.

Pretending we were ignorant about the aviation limitations that penguins exhibit, the argument is strong, on account that if the premises are true, the conclusion is more than likely to be true. Knowing that penguins, in fact, cannot fly shows that categorising (non-guaranteeing) arguments as valid, invalid, strong or not, does not change the reality of the outcome. In this case, more information is required to achieve accuracy with reality, such as "Penguins have never been seen to fly". When information about exceptions such as this are made available, they broaden the information available to form the argument. Engineering information used in decision-making can also be arranged like this.

There are many ways to present an argument; a formalisation known as the Toulmin argument structure is chosen here for its format in application and its previous known application to aircraft design methodology research for verification and validation of design choices. The choice of application to aircraft maintenance for use by maintenance managers is through informal logic, as opposed to more strict conventions demanded by formal logic. The reasons for this stem from the context of use and the expertise of the target users, where informal logic is seen to be a more appropriate representation of everyday argument style. It is intended that this framework will enhance situational deliberation, unambiguous decision support, and risk communication through careful structuring of information and sound reasoning.

Toulmin arguments begin with the *[GROUNDS]*, the base facts or starting point of the argument, and end in a *[CLAIM]*, the conclusion that the user wishes to present as a convincing argument. In our framework, the *[GROUNDS]* are the information with which the user has to go on, and the *[CLAIM]* is either that the aircraft is safe or unsafe. Examples of *[GROUNDS]* are:

- The crack in aluminium frame X5459 is 6mm long, radiating from a fastener hole.
- Technical Publication System 1/subsystem 2/maintenance/repair manual indicates that cracks longer than 5mm are not permitted.
- Advisor A has inspected the damage and states that the issue will "not be a problem for the next sortie".

[GROUNDS] should be unambiguous to any bystander, who will be able to identify the starting point for the argument. However, to transition from "the crack in the frame is 6mm long" to "aircraft is safe to fly" requires a bridging step, an inference licence that allows the user to indicate clearly why a crack in the frame is "safe", or the equivalent of this i.e. not something that will cause the aircraft to be "unsafe". *[BRIDGE]* statements, and any relevant *[SUPPORT]* are required in order to inform reviewers of the conditional reasons or beliefs that the *[CLAIM]* should be considered to be true. The *[BRIDGE]* explicitly enables the *[GROUNDS]* to be linked to the *[CLAIM]*, with any generalisable or categorical information that providing backing for the *[BRIDGE]* also stated, in order to *[SUPPORT]* the *[BRIDGE]*. For example:

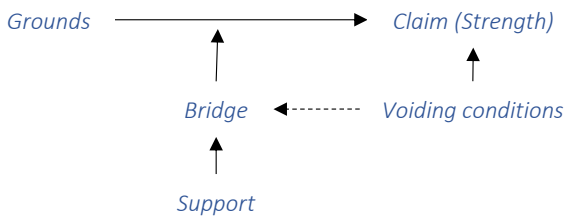
> *[GROUNDS]*: the crack in the frame is 6mm long, radiating from a fastener hole.
>
> *[BRIDGE]*: A previous crack on the same frame in a different location was cleared by the OEM to be permissible up to 30mm.
>
> *[SUPPORT]*: aluminium used in this aircraft is known to have good crack growth resilience with long critical crack lengths.
>
> *[CLAIM]*: the aircraft is safe for the next sortie.

These statements are individualistic, meaning that they are the perspective of the decision-maker. However, the user may be uncomfortable with providing a universally applicable inference in the form as above, so they may need to indicate the assumptions that the *[BRIDGE]* is conditional upon. For example:

> *[VOIDING CONDITIONS]*: Unless the loading pattern is significantly different between the other cleared crack and this damage; unless the crack has been measured incorrectly and is actually larger than indicated; unless the OEM approval for the longer crack was based upon assumptions that are invalid in this case.

Clearly, this argument has a number of frailties that make it's validity questionable (there might be disagreement regarding aluminium having a long critical crack length, or that an inference from another damage assessment is appropriate). In formal (deductive) logic, such as the arguments presented about penguins, validity cannot be claimed where the premises (the *[GROUNDS]*, *[BRIDGE]* or *[SUPPORT]*) may be true, but the *[CLAIM]* is false. In informal logic, which is more readily applicable to real-world situations, the strength of an argument is dependent upon two properties: the validity of its premises, and the strength of the inference from the premises to the *[CLAIM]*. Identifying sound informal arguments will be discussed when we introduce the heuristic decision tree.

Visually, the elements in the argument can be shown to interact as follows:



In building an argument, there are common modes that can provide a guide in arranging information and knowledge, acronymised as GASCAP:

- **(G) Argument by Generalisation.** Samples provide a limited perspective of a population, but may provide sufficient information in which to lay a convincing argument by generalisation. For example, instances of a particular aircraft fault on several aircraft may be generalizable to the remaining fleet, provided certain relevance conditions are met (configuration similarities, sufficient evidence to give grounds for correlation).
- **(A) Argument from Authority.** Some information sources may be perceived as authoritative or expert, to the extent that their word is sufficient grounds for an argument. For instance, advice from an expert technical advisor may be deemed as such, provided that there were no reasons to suspect fallacy.
- **(S) Argument by Sign.** Situations that prompt a conclusion given a particular indicator are arguments by sign. For instance, bubbling surface finish is an indicator of material build up under the paint caused by material products of corrosion, which combined with knowledge of the subject component can establish an argument for the safety of flight.
- **(C) Argument by Cause.** A causal link can be established if two factors (X, Y) nearly always appear to together. If factor X occurs, a causal argument that Y is expected to occur would be submitted on the understanding of the causal link between X and Y. However, this argument form can be fallaciously applied if there is an unrecognised third factor (Z) – correlation does not mean causation. For example, dent damage on a number of aircraft (X) might be found on aircraft with high flying hours (Y) relative to the remaining fleet, but the true cause might be the result of accidental damage during embodiment of a modification (Z)that was introduced on high flying hour aircraft.
- **(A) Argument by Analogy.** Analogy-based arguments are fundamental in law and in decisions regarding future policies, by using two observably alike situations across a number of features that can be inferred to be alike in additional features. For instance, a high life aircraft found with cracking in a particular location may provide an analogous example for other aircraft with relevantly similar life metrics.
- **(P) Argument by Principle.** A principle is a generally regard truth or property, which is applicable to a given situation of argument. In aviation maintenance, this might concern abiding by technical publications and regulations, which are generally considered to be true. Where the argument may be challenged in evaluation concerns the applicability and exceptions, the truth of the principle on which the argument relies, or whether there are other general principles that may have bearing on the argument basis.

These strategies are not exhaustive but they form a common reference set, although more specific strategies might be applied by users in their own cases. Each has conditions that may void the argument, given the nature of the strategy and it is important that the arguer considers potential pitfalls in the argument in order to make the argument strong.

The final aspect in the argument formation used here is a measure of the *[STRENGTH]* for an argument. Linguistic in nature, measures of argument strength must be carefully considered in order to appropriately convey the arguers' true belief about the strength of an argument. Noting how sensitive subjective boundaries are to individual perceptions, only three categories are listed to provide the criteria to measure the argument against: "strong" support for the *[CLAIM]*, "some" support for the *[CLAIM]*, and "no" support for the *[CLAIM]*:

- **STRONG:**
  - A LINK BETWEEN THE CLAUSES (GROUNDS, BRIDGE AND SUPPORT, CLAIM) IS KNOWN TO EXIST AND IS APPLICABLE TO THE SITUATION BEING CONSIDERED. THERE IS A RELEVANT CAUSAL CHAIN OR CORRELATION THAT IS ACCEPTED BY CONSENSUS OR AUTHORITY.
- **SOME:**
  - THE LINK IS NOT STRONG, I.E. A LINK BETWEEN THE CLAUSES IS POSSIBLE, BUT CORRELATION OR CAUSALITY IS NOT ACCEPTED BY CONSENSUS OR AUTHORITY, OR MAY BE IRRELEVANT TO THIS SITUATION.
- **NO:**
  - THERE IS NO CLEAR LINK BETWEEN THE CLAUSES ABLE TO BE FORMED FROM AVAILABLE INFORMATION, KNOWLEDGE OR DATA.

The definitions attend to the user's understanding of their argument. This is not a qualitative measure of the risk probability, but a subjective and qualitative measure of argument strength. Argument validity is the formal measure of an argument's completeness, but time-constrained maintenance situations infer arguments of a non-deductive nature, meaning that the truth of the premises makes the *[CLAIM]* more (or less) probable, rather than guaranteed. The importance of this step acknowledges that in situations of poor information or lacking authoritative information, the absolute truth of the *[CLAIM]* cannot be guaranteed. Without accounting for the information supporting a *[CLAIM]*, a decision-maker may be misled in their belief that a *[CLAIM]* is more certain, confirmed, or trustworthy than it otherwise is. Clear identification of the *[GROUNDS]* and supporting ideas for a *[CLAIM]* also help transparent scrutiny of the argument and assist in challenging biases.

The presentation of information in this manner is only appropriate because it is being applied in a situation where a judgement-based decision is *compelled*. If more time, resources or information permits more thorough and optimised assessment of the situation it should be followed. To aid users in identifying when to apply argument to support judgement, a decision tree is required.

Heuristic decision tree

Arranging information into an argument form does not alone constitute an appropriate solution to the issue of applying judgement to decision-making. Similarly, defining objective criteria for permitting flying or not, via a decision tree or other rule-based definition, does not enable situational flexibility and may waste resources where an individual sees room to make a valid argument using their judgement. A hybrid approach is thus appropriate for justifying the use of judgement for making a decision when unsupported by directly relevant decision support material. Although an objective criteria such as "if no authoritative decision-support information is available, use judgement" dictates a clear threshold, it is too specific and would encourage long checklists of situations where judgement-use was permitted and prohibited. Innumerable context and task factors have bearing on the need to use judgement in engineering maintenance decision-making. It is appropriate to challenge whether judgement-use in a given situation is, or was (if reviewing post hoc), necessary, so being able to reflect upon the situation and the particular need for using judgement is a primary consideration. In addition, it is appropriate to assume that individuals alone will make fallacious arguments on account of inherent cognitive and motivational biases. The principles that enable the use of judgement for decision-making in operational situations are thus summarised:

- **NECESSITY:**
  - MUST JUDGEMENT BE APPLIED FOR THIS PROBLEM IN THIS CONTEXT?
- **SOUNDNESS OF CONSTRUCTION:**
  - DOES THE INFORMATION AND KNOWLEDGE AVAILABLE ENABLE A SOUND ARGUMENT TO BE CONSTRUCTED?
- **SCRUTINY AND CONSENSUS:**
  - IS A CONSENSUS POSSIBLE FOR THE CLAIM, IT'S SUPPORTING PREMISES AND CONDITIONAL ASSUMPTIONS?
- **TIME-CONDITIONALITY:**
  - JUDGEMENT-BASED ARGUMENTS ARE CONDITIONAL ON THE INFORMATION AND RESOURCES AVAILABLE AT A GIVEN SPACE AND TIME; A CHANGE IN THE STATE OF INFORMATION OR RESOURCE THAT CHANGES THE PREMISES OF THE ARGUMENT RENDERS EXISTING ARGUMENT VALIDITY AS VOID.

The first three principles are arranged into three "gates", through which the user must be able to answer "yes" to substantiate a judgement-based decision to fly. Saying "no" at any gate returns the user to a decision to default to authoritative instructions, ground the aircraft, or seek further advice. The final principle regarding time-conditionality is a holistic consideration that has bearing on the wider management of a given fault, which is not directly addressed in this training.

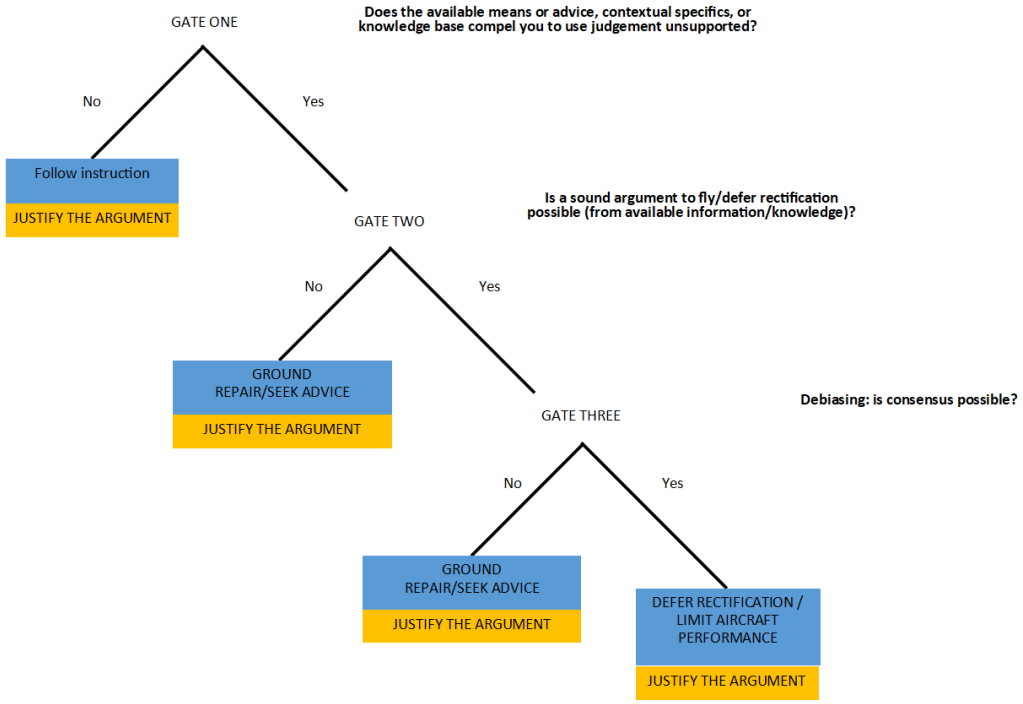


Figure 1 – heuristic decision tree for judgement-based maintenance decision-making.

Gate 1

Gate 1 considers whether the situation compels the user to use judgement, accounting for the time, information, resources and cognition available in the current context of the decision. For instance, there might be no technical data available for a particular fault, meaning there is no *means* to

determine whether the fault is permitted to be deferred or not. Alternatively, technical information may be available but given the operational necessity (the *contextual specifics*) of the situation, the user may form an argument around the technical information being irrelevant to the situation. The situation and user within the situation determines whether judgement is necessary. Users may find it helpful to record the argument for why they are not compelled to use judgement, if they expect to be challenged on these grounds by others with alternative perspectives. This is represented by the dotted line in Figure 2.

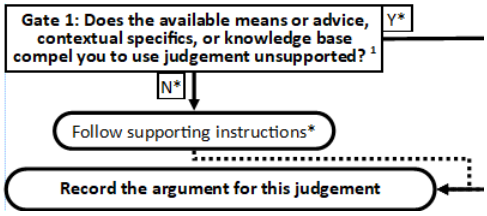


Figure 2 – Gate One of heuristic decision tree

Gate 2

Gate 2 requires consideration of whether a sound (or cogent) argument can be formed from the available knowledge and information. A cogent argument requires that:

- **IT IS ACCEPTABLE TO THOSE RECEIVING THE ARGUMENT (A).**
- **THE PREMISES ARE RELEVANT TO THE CONCLUSION (R).**
- **IT HAS SUFFICIENT OR GOOD GROUNDS FOR THE CONCLUSION (G).**

These conditions are conveniently memorable by noting that ARG are the first three letters of "argument". However, a valid argument must also address alternate perspectives, so called dialectical requirements. These may be addressed in two ways: first, by providing appropriate *[VOIDING CONDITIONS]*, containing any assumptions that may be broken or uncertainties that are not bounded, which should be clearly described to enable the transparency of the argument to be accepted; and second by using the final gate to invite scrutiny and alternative positions. As was the case for Gate One, users may find it useful to record the argument for why it is not possible to make an argument that supports fly/defer rectification, in order to placate stakeholders with differing opinions on the situation.

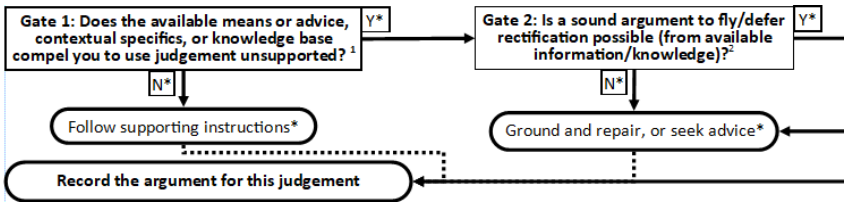


Figure 3 – Gates One and Two of the heuristic decision tree

Gate 3

Gate 3 addresses the potential for individual bias and fallacy, and incorporates alternate views into the argument. These factors may have bearing upon the accuracy of the argument premises or conclusion unforeseen by the decision-maker's perception. The dialectical consideration requires that the argument sufficiently caters for alternate positions by asking for consensus from available peers. It is not unreasonable to assume that maintenance managers are surrounded by similarly qualified individuals, team members or colleagues who can provide a cross-check of judgements. If the positions asserted by these individuals are not adequately accounted for (and are themselves considered as reasonable premises), perhaps as exceptions or assumptions, then the argument is not cogent and should be considered as unsafe.

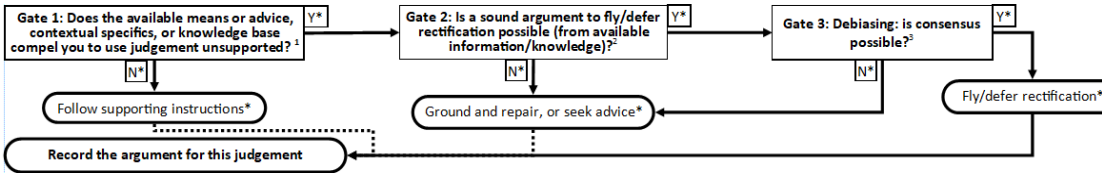


Figure 4 – Gates One, Two and Three of the heuristic decision tree

Heuristic decision making is widely known to be influenced by biases and fallacies. Including second opinions formally can reduce the opportunity for cognitive and motivational biases to influence the application of judgement to a situation. Thus, seeking consensus to the described argument premises and conclusion can add weight to the argument's claim of sufficiency. If consensus cannot be achieved, the argument may be formed upon unchecked biases; when handling safety decisions, this is principally inferior to making attempts to address the potential for fallacy.

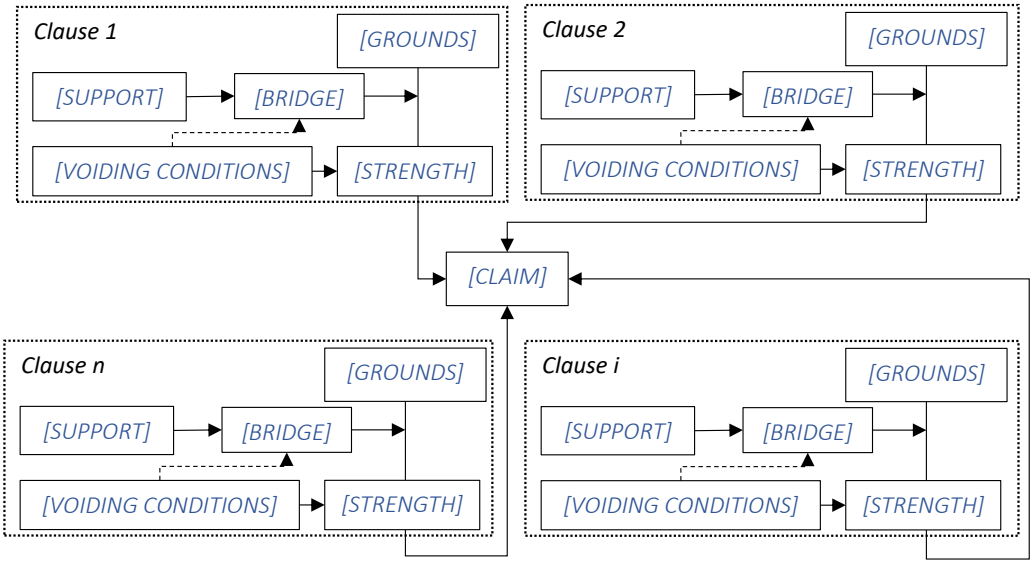
How does one know if the argument is sufficient, even with consensus? Decision-makers are limited by their cognition, knowledge and access to information. Operationally constrained situations present a closed world view – the state of information will not improve beyond the information they possess themselves or have access to, and the decision will be made under these constraints of the situation. Sufficiency, in terms of the argument completeness in an operational situations, must serve two purposes: that the argument is sufficient to be cogent and to justify the course of action instructed by the *[CLAIM]*; and that the argument is sufficiently developed within the time constraints of the context. An argument that claims to be rational under such conditionally-dependent and situational factors is seen to be boundedly rational. The rational choice made is satisfactory (but not necessarily optimised) given the situational constraints. If unsupportable under the principle of bounded rationality, the argument should be rejected on the basis that to continue flight is irrational given the current information and knowledge available. Recall that the situation predicated in the use of the heuristic decision tree: a judgement-based decision is compelled.

Multiple Grounds

An argument in support of "fly" is likely to necessitate justification of more than one ground, drawn from different available information. For instance, given a fault of bearing case cracking on a load bearing moveable component, one should expect to address and justify several aspects of the issue relative to a favourable argument to fly on the next sortie, such as:

- Further deterioration of the crack on the next sortie.
- The function of the bearing within the component on the next sortie, or in general.
- The consequences of failure if the crack propagated to failure on the next sortie.

Visually, this may be represented as follows, where *n* arguments contribute to the same *[CLAIM]*:



Recording information selection

Failure sequence is not the only aspect to consider in how a fault may influence airworthiness. Wider systems implications (air and ground human elements, functional aspects, organisational reputation aspects) may be considerations for decision-makers. Once familiar with the terms, users will be able to efficiently identify the *[GROUNDS]* that they will link soundly to a conclusion. In extremely time-limited situations, experienced decision-makers will intuitively prioritise the information that is necessary to act upon, which may only be a fraction of the available information. Having these such situations documented provides an explicit record of the whole judgement and can enable both post-hoc review and training opportunity for novices to learn the cues that more experienced practitioners act upon. Clearly such situations may be influenced by bias, but the decision-making conditions are non-optimal, so the solution should also be expected to be non-optimal. This works for both potential decision outcomes: to elect to ground the aircraft, or to elect to fly the aircraft. Consequently, depending upon the contextual circumstances, users may find it useful to document the argument if stakeholders are critical of the justification for answering “no” to gates 1 and 2.

PRACTICE

To support the use of the framework, a template form is provided at Annex A, as both a prompt and as a layout to populate if this helps structure consideration of the situation.

Process for using the framework

The framework is flexible to the situation and user, meaning that completion of each step is satisfied by the individual within the perceived limits of the situation. The process will be described sequentially.

Gate 1 – *"does the available means or advice, contextual specifics, or knowledge base compel you to use judgement unsupported?"* - assess the available information, knowledge and context to determine whether it is necessary to proceed with a judgement-based decision. Informally, this step accounts for the subjective decision to make a decision based upon judgement. Are you required to use judgement? This may be obvious (technical information might not exist for the component or fault that is being assessed), or may not be as clear cut as first imagined, for instance if the technical information appears to be inadequate or irrelevant to the situation being faced. The key is whether the situation is considered to be supported (where information is immediately usable to support the situation) or unsupported (where information is unavailable within the decision horizon). An individual may consider their position unsupported if they have grounds to question the relevance or accuracy of supporting material. If the decision-maker considers the situation to be supported, then the supporting information should be followed as directed. Otherwise, step 2 should be actioned.

Gate 2 – *"is a sound argument to fly/defer rectification possible (from available information/knowledge)?"* - does the information I have access to enable me to form a valid argument to support a decision to permit the aircraft to fly? There are discrete sub-steps that form an iterative process for building an argument:

2.1 – Recognise the extent/boundaries of the argument with respect to the [CLAIM] and [GROUNDS].

[CLAIM] – to what extent does the *[CLAIM]* need to made? What are the properties of the next flight, or flights, if the claim is being made to cover a longer period of operation.

[GROUNDS] - what information is available to be considered within the decision horizon?

2.2 – [GROUNDS]. Select the useful and relevant information that will be used as *[GROUNDS]* for the argument. Example *[GROUNDS]* may be:

EXAMPLE [GROUNDS]

- *Documentation*
- *Observable fault dimensions*
- *Known history of fault (on subject aircraft and fleet)*
- *Known aircraft configuration*
- *Known design data (including incomplete data)*
- *Known aircraft usage*

2.3 – [BRIDGE]. Construct the bridging statements (from available information, data or knowledge). Example *[BRIDGE]* information may be:

EXAMPLE [BRIDGE]

- Aircraft usage (single aircraft or fleet)
- Maintenance history (known, inferred, advised about, recalled)
- A mitigation option/strategy that make conditions for the claim viable
- Design data (if actionable and perceived as relevant to the issue)
- Individual knowledge awareness (known as meta-cognition) – knowledge of the situation/aspects involved.
- Consequence visualisation.

2.4 [SUPPORT]. [SUPPORT] information (generalisations and categorical information that relates to the *[BRIDGE]*) may be formed from information such as:

EXAMPLE [SUPPORT]

- Justification/necessity for mitigation (may be framed by resources available/context).
- Trustworthiness of information source (design data, advisor, textual information etc.).
- Relevance of information (to the situation).
- General principles, regulatory guidance or corporate accepted practices.

2.5 *[VOIDING CONDITIONS]* are the conditions that would invalidate the constructed argument. If voiding conditions become true, they would cause the *[BRIDGE]* and *[SUPPORT]* to become questioned and hence the *[CLAIM]* would not be valid.

Continuing the examples from the *[GROUNDS]*, assuming that they're concerning a fault identified on an aircraft in a fleet, these three links (*[BRIDGE], [SUPPORT]* and *[VOIDING CONDITIONS]*) may be formed as follows:

- *Documentation* – "technical publications indicate the fault is out of limits".
  - *[BRIDGE]*: technical publications are authoritative documents.
  - *[SUPPORT]*: publications are written by specialists and supported by design data.
  - *[VOIDING CONDITIONS]*: if the publication is generic in application to the situation; if interpretation of the limits is ambiguous to the situation; if publication may have been written without the current context considered.
- *Known aircraft fault data* – "3 other aircraft in fleet have a very similar fault indication".
  - *[BRIDGE]*: other fault data is sufficiently representative of current issue.
  - *[SUPPORT]*: systemic faults caused by normal loading will occur on multiple aircraft.
  - *[VOIDING CONDITIONS]*: conditions of fault arising have different causal chains; if fault is not systemic (i.e. unique).
- *Known aircraft usage data (landings fatigue index, flying hours, pressurisations, etc.)* – "high total landings relative to remaining fleet".
  - *[BRIDGE]*: aircraft is the fleet leader in total landings, and landing loads are believed to be a causal factor.
  - *[SUPPORT]*: Fleet leaders will exhibit unforeseen faults before younger aircraft.
  - *[VOIDING CONDITIONS]*: fault is not driven by landings.

- *Knowledge about general principles* – "corrosion found on aircraft skin material around fastener head".
    - *[BRIDGE]*: corrosion reduces material thickness but visible cracking will precede failure in skin; single fastener failure will not jeopardise structural integrity of panel.
    - *[SUPPORT]*: surface corrosion changes material properties and weakens the parent material effective strength; fastener corrosion is common.
    - *[VOIDING CONDITIONS]*: localised effects of corrosion are more severe than perceived with greater material thickness reduction, or cracking present; fastener failure is critical to structural integrity of skin section.

Depending upon the nature of the situation, the *[GROUNDS]*, *[BRIDGE]*, *[SUPPORT]* and *[VOIDING CONDITIONS]* may be perceived as contributory to an argument to fly, or as incompatible.

2.6 – [CLAIM] and [STRENGTH]. Formulate the conclusion that the *[GROUNDS]*, *[BRIDGE]*, *[SUPPORT]* and *[VOIDING CONDITIONS]* instruct. Forming the conclusion any earlier in the process may result in warrants or backing being framed in a bias manner toward the conclusion, rather than being treated impartially. For situations based upon subjective support information (for example if assessing the severity of corrosion damage to a plate, or the potential consequences of a failed non-structurally-load bearing fairing), this could influence the way in which the faults are deliberated upon. In other words, it prevents one from confirming their instinctive opinion. In addition, a broad categorising statement reflecting the strength of each argument clause in supporting the conclusion should be given, using the nominal scale to illustrate "strong", "some", or "no" support for the conclusion. The lowest common denominator provides the main indicator of the overall strength of the argument. These help direct attention to those clauses that could be prioritised with improved information or analysis.

- **Strong:** a link between the clauses (*[GROUNDS]*, *[BRIDGE]* and *[SUPPORT]*, *[CLAIM]*) is known to exist and is applicable to the situation being considered. There is a relevant causal chain or correlation that is accepted by consensus or authority.
- **Some:** the link is not strong, i.e. a link between the clauses is possible, but correlation or causality is not accepted by consensus or authority, or may be irrelevant to this situation.
- **No:** there is no clear link between the clauses able to be formed from available information, knowledge or data.

[STRENGTH]

2.5 – Soundness. Review the complete argument for soundness from informal logic using ARG (acceptable, relevant, good/sufficient *[GROUNDS]*).

- *Acceptability*: it is acceptable to those receiving the argument (A).
- *Relevance*: the premises are relevant to the conclusion (R) and all relevant information has been taken into consideration.
- *Good grounds*: It has sufficient or good grounds for the conclusion (G).

ARG

If the argument formed from the available information is not valid, the process terminates in grounding the aircraft, repairing the fault, and/or seeking further advice. Otherwise, step 3 should be actioned.

Gate 3 – "Is consensus possible?" - Addressing the influence of bias by gathering consensus on argument. Third party perspectives enrich individualistic judgement by helping to nullify sources of bias through alternate perspectives (assuming the judgement is from a different perspective). An argument is strengthened by being able to account for alternate perspective and by enabling independent reviewers to assess the argument, potential sources of fallacies can be more readily addressed. Fallacy occurs when one produces an argument that contains at least one serious and identifiable defect, i.e. there is a borderline that has been crossed between the argument being acceptable to one person, and unacceptable to another for an identifiable reason. Approaching fallacy pragmatically is important in time-critical and resource-limited situations, as formal analysis of argument dialogue or rhetoric is not likely to be useful or meaningful when under time pressure. Cross-checking an individual's argument basis provides time-efficient opportunity to expose problems in the assessment by challenging the validity of the argument independently. Although not an optimised process for debiasing beliefs, it provides a situationally appropriate means to mitigate the effects of fallacy on validity of an argument. The ARG test for cogency (above) provides the handrail for independent scrutiny of the argument in order to agree consensus.

Consensus at gate 3 provides a limited but independent validation of the argument. If consensus is gained, the argument is valid and debiased as far as is practicable under the constraints of the situation. This means that there is a sound argument in support of flying the aircraft under the faulty condition. If consensus is not possible and valid differences of opinion cannot be incorporated in support of the *[CLAIM]* (or, reciprocally, the *[CLAIM]* cannot be adjusted) then the decision termination is to ground the aircraft, repair the component, or seek further advice.

Arguments constructed on information available at a given point in time are time-conditional, meaning that new information received at a later time may enable the argument to be revised, as indicated by the fourth principle above.

An example case now describes the application of the framework to a realistic decision-making situation.

Case study #1 – Space Shuttle Challenger

Context: shortly after take-off, the Space Shuttle Challenger was lost in 1986 in a catastrophic accident that killed all crew, when an O-Ring seal failed to contain hot gases in the solid rocket boosters (SRB) during take-off. The temperatures on the launch morning were predicted to be far colder than any previous launch. This alarmed technical personnel into considering the data they had about the ability for the O-ring seals on the solid rocket boosters to remain sufficiently ductile in order to seal the joints between the rocket booster sections and contain the hot gases during launch. A debate ensued between contractors and government authority personnel, which highlighted conflicting data: between inconsistent real launch data, which showed evidence of seal erosion (caused by hot gases leaking) across warmer and cooler launches; inconsistent behaviour from the rocket booster contractor who had never previously wanted to stop a launch because of the cold effects on the O-ring seals; and sub-scale component testing data under simulated cold conditions that indicated the performance of the O-ring was sustained. However there was no complete system data to support performance at cold temperatures. Consequently, the decision to launch was based on engineering judgement:

> *"Most of the concerns that we had presented were qualitative in nature. At that particular time we had a very difficult time having enough engineering data to quantify the effects that we had been talking about. A lot of it was based on 'engineering feel' " - Morton Thiokol's Jack Kapp (p308, Vaughan (1996)).*

A post-hoc review of the situation using argument is applied to illustrate the types of information that may have been arranged to demonstrate the lack of validity in the argument to continue to fly.

Gate 1: "does the available means or advice, contextual specifics, or knowledge base compel you to use judgement unsupported?" The quotation indicates that engineers lacked data in which to form an analytical solution to whether the cold temperature would impact the function of the O-ring seal ("we had a very difficult time having enough engineering data to quantify the effects that we had been talking about"), requiring instead to rely on 'engineering feel'. This is evidence that members of the decision-making unit were compelled to use their judgement.

Gate 2: "is a sound argument to fly/defer rectification possible (from available information/knowledge)?" To form a *[CLAIM]* about whether to launch or not, sub-arguments may have been formed on the following *[GROUNDS]*:

1. *Forecast ambient temperature at launch time is below data base for whole system*
2. *SRB sealing resilience given sub-scale component level and erosion resistance tests*

Taking these *[GROUNDS]* to form sub-arguments individually:

1. *[GROUNDS]*: *Forecast ambient temperature at launch time is below data base for whole system*
    - *[BRIDGE]*: No evidence that the full-scale vehicle has previously launched safely in forecast temperatures; launch evidence is not correlated (erosion seen on normal and low temperature launches).
    - *[SUPPORT]*: There are no clear temperature limits established for the launch vehicle, despite heavy erosion being identified following launch at 53°F. No data consistency proving full-size system safety at cold temperatures. Forecast temperature is lower than all previous launches.

- *[VOIDING CONDITIONS]*: If SRBs have been designed using inherent temperature limits common in all NASA ballistic missile design that contain the forecast temperature range.
- *[STRENGTH]*: "No"

Is this sub-argument cogent? This argument appeals to an *argument from sign*, whereby the temperature is the indicator of safe launch. Without temperature limits for the system (backing), the arguers are basing their *[BRIDGE]* upon the cold temperature being an indicator that they are operating the system outside of whole system known performance. The *[STRENGTH]* is "no" because it was apparent from conflicting data that they had no correlation between the occurrence of damage to the O-ring seals on live launches, and component test data did not simulate launch dynamic forces, like wind shear effects and bending of the SRB during launch. Therefore the sub-argument is not cogent because there isn't a sufficient data base ([G] from the ARG test).

2. *[GROUNDS]*: ***SRB sealing resilience given sub-scale component level and erosion resistance tests***
    - *[BRIDGE]*: Sub-scale tests provide indication of resiliency to temperature and erosion; sub-scale tests but do not simulate the launch forces on joint rotation and flexing.
    - *[SUPPORT]*: Failure modes that combine seal hardening (owing to low temperature), erosion and joint rotation/flexing are not simulated by sub-scale tests.
    - *[VOIDING CONDITIONS]*: Failure mode combining hardening, erosion and structural flexing are considered insignificant; sub-scale tests are shown to cater for these combined failure mechanisms. Prevailing weather conditions do not lead to seal hardening or joint rotation/flexing.
    - *[STRENGTH]:* "Some"

    Is this sub-argument cogent? This argument is being posited as an *argument of analogy* between the component tests and the full-scale vehicle operation. If the tests are satisfactory at cold temperatures, the full-scale vehicle can be inferred to be satisfactory at cold temperatures. Therefore, under some conditions the argument may give support to the decision to fly. However, the analogy is not valid as there are exceptions to the component tests, namely that the full-scale vehicle is subjected to other dynamics during live launch that were not tested during the cold testing of the component. Therefore, the argument is not valid because it is neither acceptable [A] or sufficient [G]. This decision should terminate at Gate 2, seek further advice to assert the validity of the argument with new information.

Gate 3: "Is consensus possible?" As a demonstration of the mechanics of Gate 3, we explore Gate 3 as a demonstration despite the decision being terminated at Gate 2.

There is evidence from technicians who were present during discussions that they did not agree with the recommendation to launch. Roger Boisjoly believed the Shuttle project wouldn't be expected to launch in anything less than 53°F (previous lowest launch temperature, which resulted in O-ring seal erosion) and that "in launching below the data base they were moving away from goodness" (p317, Vaughan (1996)). Jack Kapp believed that there was no quantifiable data that indicated that the complete Shuttle vehicle was safe at the forecast temperature (p308, Vaughan (1996)). Robert Ebeling believed the planned launch was below the data base of launch temperatures (p87, Rogers (1986)).

Each of these positions refer to sub-arguments 1 and 2, indicating that either they didn't believe the arguments were valid (owing to either sufficiency of data, or relevance of the data they had access

to), or they would not provide consensus on the argument to fly. Using the framework, the Shuttle would not have launched, either at Gate 2 (in assessing the soundness of the argument), or Gate 3 in requiring consensus.

This case is a post-hoc review, so testimonies may suffer hindsight or confirmation bias, however cross-referenced accounts of the actions that were taken are sufficient to indicate that in the face of judgements that opposed the argument to fly, the judgements were disregarded.

Case Study #2 – Cargo Aircraft Damaged Undercarriage Door (fictional)

Context: a multi-engine cargo aircraft has been in operation for 12 years, with a fleet of 24 aircraft. All aircraft operate similar flying profiles and have flying hour accruals between 4500 and 9000 flying hours, full stop landings of between 500 and 2500, and total landings between 550 and 3500. The aircraft are currently deployed worldwide supporting multiple operations, although there is usually only one aircraft supporting each operational location.

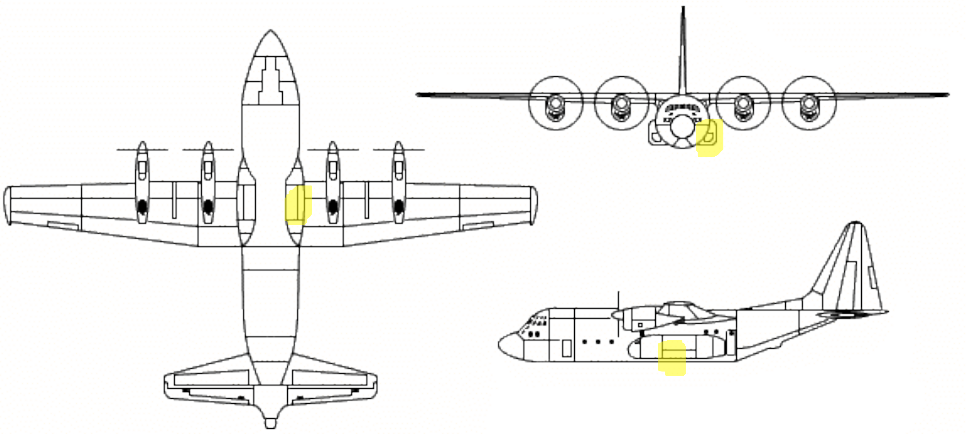


Figure 5 – multiengine aircraft arrangement, with damage location indicated with yellow highlight. This Photo by Unknown Author is licensed under CC BY-NC-ND.



Figure 6 – main landing gear with damage location marked with red circle. This Photo by Unknown Author is licensed under CC BY-SA



Figure 7 - Inside of main landing gear door, illustrating dent damage.

Fault: *A request for advice has been received from a frontline squadron operating in support of an overseas operation as x4 dents have been found on the inside of the starboard side main landing gear door. The subject aircraft has a relatively high number of full stop landings and is mid-fleet in flying hour usage and total landings. The aircraft is currently operating with mid-level cargo transit flights between regional bases. No limits have been provided for the door structure in particular, but as the panel is a sandwich construction with aluminium skins and a honeycomb core, a generic section of the technical publications indicates that the damage is outside of generic limits given for metallic honeycomb structures, owing to the number of dents and their proximity to one another. No spare main landing gear doors are available in the aircraft's current location; though there are 2 in stock at main base and to ship one would take 10 days. A trusted technician in your team advises that the maintenance procedure for replacement of the door is understood to take 6 hours. A review of faults reported in the last 3 months indicates that similar damage has been accrued on 3 other aircraft in the same location as the most recent request for advice. The previous occurrences were all deferred by the operating Squadrons until the next scheduled maintenance period, with mitigation detailed as "after flight checks to monitor the dents for deterioration". A conversation with the OEM/manufacturer structural specialist engineer has suggested that "the limits in the technical publications should be followed" but they inform you that the door carries flight loads as a component of the semi-monocoque construction of the aircraft fuselage. Formal advice from the OEM requires detailed fracture analysis which would take at least 14 days. You are required to determine whether the subject aircraft should continue to operate. Assume that your colleague is available and has provided agreeable testimony to your assessment.*

Gate 1: Is the use of judgement justified? The aircraft is in demand to support operations with no resilient spare aircraft. Although there is supporting documentation for this damage, the operating evidence suggests that a number of aircraft have been flying with the aircraft without issue, which undermines the need for immediate recovery action dictated by the technical publication. Therefore,

the dependence on supporting information suggests that it is justifiable to consider the situation using judgement.

Gate 2: Is a sound argument to fly the aircraft possible? The primary *[GROUNDS]* that influence personal judgement on the situation are identified from the case information to form the argument:

1. Dents are out of limits in accordance with generic technical publication.
2. Other aircraft in the fleet with similar damage to the same component have continued flying with no deterioration to the damage.
3. OEM structural specialist advises that the technical publication limits should be followed.
4. Panel is a sandwich construction with aluminium skins and honeycomb core.

Taking each of these in turn to form an individualistic argument:

1. *[GROUNDS]*: *Dents to main landing gear door inner surface are out of limits in accordance with generic technical publication.*
    - *[BRIDGE]*: generic technical publication are believed to have been written to cater for the limits of the flying envelope.
    - *[SUPPORT]*: technical publications are authoritative documents; technical publications do not cover every fault possibility.
    - *[VOIDING CONDITIONS]*: assuming the aircraft is not going to be operating at the limits of the flight envelope for the undercarriage door; assuming the generic limits in the technical publication are not directly relevant to the main landing gear door inner surface.
    - *[STRENGTH]*: "some"

    Is this sub-argument cogent? This sub-argument is an *argument from principle*, that the technical publication limits are generic and therefore not specifically catering to the loading and functional aspects of the door. It is generally known that technical publications do not cover all circumstances, otherwise the need to contact technical support would not be so prevalent. Therefore there is a causal link between belief in the technical publications when there isn't specific guidance, and technical publications being deficient of information. Hence the *[STRENGTH]* to defer repair of the dents is "some". The validity of this sub-argument is questionable on the basis of the "believed to have been written to cater for….." statement in the *[BRIDGE]* – this is acceptable to the arguer but may be challenged at Step 3. So for the time being, the argument is valid.

2. *[GROUNDS]*: *Other aircraft in the fleet with similar damage to the same component have continued flying with no deterioration to the damage.*
    - *[BRIDGE]*: fleet aircraft flying similar profiles; dents across the other reported incidents are similar in size but less in number.
    - *[SUPPORT]*: structural loading is sufficiently similar across flying profiles; damage propagation is a similar rate to other incidents.
    - *[VOIDING CONDITIONS]*: assuming the damage is similar in severity to this particular case; assuming the aircraft are operating similar flying profiles, assuming the deterioration is similar across damaged aircraft.
    - *[STRENGTH]*: "some"

    Is this sub-argument cogent? This sub-argument is an *argument by generalisation* based on the arisings found on other aircraft. There is correlation from the other aircraft with similar damage, however without those aircraft life data or other information to pertaining to the nature of the damage, the correlation between the damage existing and the aircraft being safe to fly may

be circumstantial. For instance, older aircraft (with older main landing gear doors) may suffer aging effects on the panel, particularly in the undercarriage bays which are regularly open to environmental forces, so if our aircraft is significantly older or with higher landing metrics, the argument may not be particularly relevant. Therefore, there is "some" support for the *[CLAIM]*, but it is conditional. These aspects are captured in the argument within the *[VOIDING CONDITIONS]* so the argument can be considered to be complete, and valid as a consequence of no reason to doubt the ARG aspects of the argument.

3. *[GROUNDS]*: *OEM structural specialist advises that the technical publication limits should be followed.*
    - *[BRIDGE]*: OEM engineer's advice is informal and analysis is not available immediately; individuals would not be expected to commit their (or their company's) reputation for ad hoc advice.
    - *[SUPPORT]*: OEM personnel will toe the line of company policy.
    - *[VOIDING CONDITIONS]*: unless OEM engineer's advice was accurate regarding the main landing gear door and not a generalised conservatism; unless assumptions about the OEM engineer's motivation are not accurate.
    - *[STRENGTH]*: "some"

Is this sub-argument cogent? This sub-argument is an *argument by authority*, but one in which the authoritative position is taken to be a fallacy by it not being independent of organisational motivations (reputational, policy or legal considerations). It is therefore a negatory argument, i.e. that in this case, the authority should not be believed because of given reasons. However, knowing that OEM advice is normally authoritative, the *[BRIDGE]* does not strongly support the *[CLAIM]* because else the OEM would not be credible. Therefore, the *[STRENGTH]* is "some", conditional on the situational factors – time limit on OEM advice, OEM advice is informal and not supported by analysis etc. These factors are addressed by the *[VOIDING CONDITIONS]*, so the argument is considered valid.

4. *[GROUNDS]*: *Panel is a sandwich construction with aluminium skins and honeycomb core which transfers fuselage skin loads once the undercarriage doors are closed.*
    - *[BRIDGE]*: dents on honeycomb structure can create disbonds between the outer skin and the honeycomb; honeycomb construction is resilient to damage; failure of the main landing gear door structure will not cause loss of structural integrity in flight as it transfers aerodynamic and fuselage loads; failure of main landing gear door may cause obstruction to deployment of the undercarriage.
    - *[SUPPORT]*: honeycomb structures are used to provide high stiffness traded off for weight; the aircraft operates with main landing gear doors open routinely, meaning total loss of the door will not result in loss of aircraft; significant deterioration of the panel to cause deformation and subsequent obstruction of the undercarriage operation has not occurred on other aircraft with damage.
    - *[VOIDING CONDITIONS]*: provided that metal skin construction does not disbond from honeycomb; provided that catastrophic disbonding of the honeycomb does not occur between inspections; provided that other aircraft occurrences are sufficiently similar.
    - *[STRENGTH]*: "some".

Is this sub-argument cogent? This sub-argument is an *argument by cause*, honeycomb structures are resilient to damage provided that the honeycomb has not disbonded. This is a particularly hazardous and arbitrary *[CLAIM]* because it hides an assumption that the potential for disbond is considered across the whole flight profile. The sub-argument continues to *[BRIDGE]* to

the "safe to fly" recommendation by considering that if the door was to fail, the resulting failure chain would not lead to loss of the aircraft, which is also supported by an *argument by generalisation* regarding the lack of failure seen on other damaged aircraft. While acceptable to the arguer, and with the assumptions contained in the *[VOIDING CONDITIONS]*, this may be contested when seeking consensus.

Subject to the argument clauses being deemed as appropriate, the *[CLAIM]* can now be established.

[CLAIM]: *there is some strength in the argument that the dents on the inside of the main landing gear door do not impact the declaration that the aircraft is "safe to fly" for the next sortie.*

Within a constrained time, an argument has been formed from four fragments of evidence, conditional upon the clause accuracy, relevance and sufficiency (ARG), and the *[VOIDING CONDITIONS]* not occurring. They are a personal indication of the evidence that constructs a valid argument in the time frame available.

Gate 3: Gather consensus on argument. Although this step is unable to be handled discretely, we are told to assume that there is consensus from a colleague.

We therefore arrive at the decision, which is to fly the aircraft subject to several assumptions and presumed conditions. Had there been any disagreement, dissent or irregularity raised from those being sought for consensus, these issues would require resolution if the decision to continue to fly was to proceed. If a new consideration had been identified, that would also have been required to be accounted and the *[CLAIM]* reassessed, for example:

5. *[GROUNDS]*: *Cause of the denting is currently unknown.*
    - *[BRIDGE]*: Causes of dents are perceived to be accidental damage, deliberate damage, complex loading, malfunction of undercarriage. Accidental and deliberate damage may have
    - *[SUPPORT]*: an unidentified cause may result in further unexpected damage.
    - *[VOIDING CONDITIONS]*: unless a cause can be identified by checking maintenance operations in the undercarriage bays, reports of undercarriage malfunctions or strange noises, standard serviceability tests indicate satisfactory operation of undercarriage.
    - *[STRENGTH]*: "no"

The *[CLAIM]* then requires amendment to the lowest denominator:

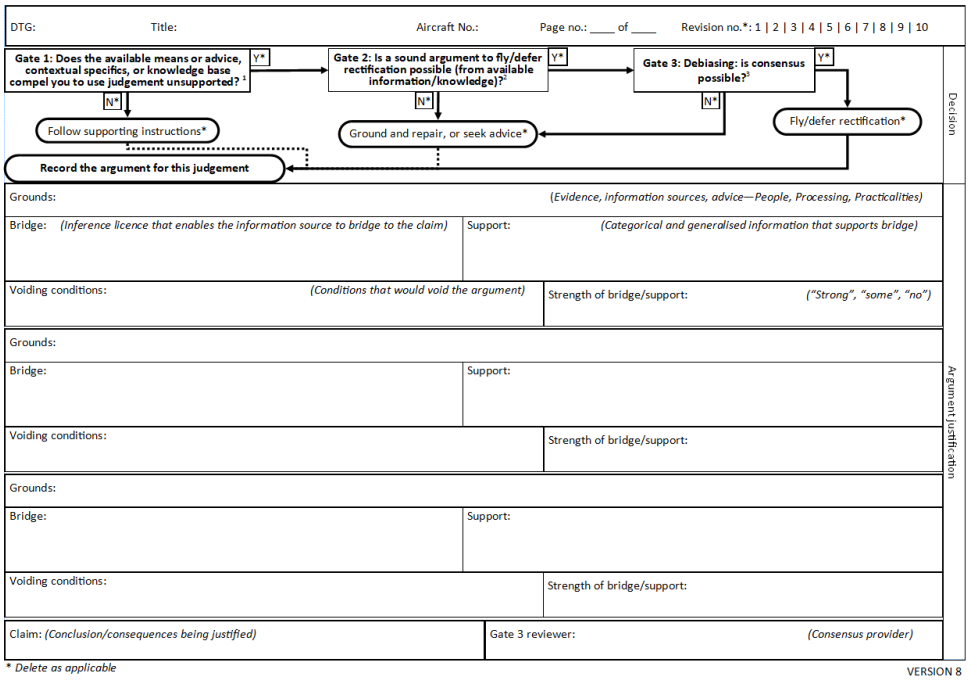
[CLAIM]: There is no strength to the argument that the dents on the inside of the main landing gear door do not impact the declaration that the aircraft is "safe to fly" for the next sortie.

The argument may be improved by gathering more information to substantiate uncertainties detailed within the *[VOIDING CONDITIONS]*.

REFERENCES

Vaughan, D. (1996). *The Challenger Launch Decision - Risky Technology, Culture, and Deviance at NASA*. Chicago: The University of Chicago Press.

Rogers, W. (1986). Space Shuttle Challenger Accident Commission Report. Available at: https://spaceflight.nasa.gov/outreach/SignificantIncidents/assets/rogers_commission_report.pdf (Accessed: 28 September 2020).


DTG:
Title:
Aircraft No.:
Page no.: ____ of ____
Revision no.*: 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10
Gate 1: Does the available means or advice, contextual specifics, or knowledge base compel you to use judgement unsupported? [1]
Y*
N*
Gate 2: Is a sound argument to fly/defer rectification possible (from available information/knowledge)?[2]
Y*
N*
Gate 3: Debiasing: is consensus possible?[3]
Y*
N*
Follow supporting instructions*
Ground and repair, or seek advice*
Fly/defer rectification*
Record the argument for this judgement
Decision
Grounds: (Evidence, information sources, advice—People, Processing, Practicalities)
Bridge: (Inference licence that enables the information source to bridge to the claim)
Support: (Categorical and generalised information that supports bridge)
Voiding conditions: (Conditions that would void the argument)
Strength of bridge/support: ("Strong", "some", "no")
Grounds:
Bridge:
Support:
Voiding conditions:
Strength of bridge/support:
Grounds:
Bridge:
Support:
Voiding conditions:
Strength of bridge/support:
Argument justification
Claim: (Conclusion/consequences being justified)
Gate 3 reviewer: (Consensus provider)
* Delete as applicable
VERSION 8

Notes supporting decision judgement tool:

Note 1. Using judgement "unsupported" means that there is/are no authoritative source(s) on which to base a decision regarding an identified fault.

Note 2. A *sound* argument can be formed when the information available to the decision-maker in a situation can be validly organised to support the claim or conclusion. A sound argument is formed when information *[GROUNDS]* is linked to a claim by a bridging licence *[BRIDGE]*, and supported by categorical or generalisable information/knowledge *[SUPPORT]*. The licences used in an argument may be context specific. A *strong* argument can be formed when a sound argument is supported by a statement of the conditions under which the argument becomes void *[VOIDING CONDITIONS]*, and by an indication of the strength of the argument *[STRENGTH]*. Arguments can be found to be sound by reviewing their:

- *Acceptability*: it is acceptable to those receiving the argument (A).
- *Relevance*: the premises are relevant to the conclusion (R) and all relevant information has been taken into consideration.
- *Good grounds*: It has sufficient or good grounds for the conclusion (G).

Note 3. Individualistically formed judgement can be biased and difficult to debias without support from experts or peers of similar repute within the domain that the argument is formed around. Consensus must be formed around the formed argument with suitably reputable peers.

VERSION 8

L.2 Quick Reference Guide

AIRCRAFT ENGINEERING MAINTENANCE

JUDGEMENT-BASED DECISION-MAKING SUPPORT FRAMEWORK

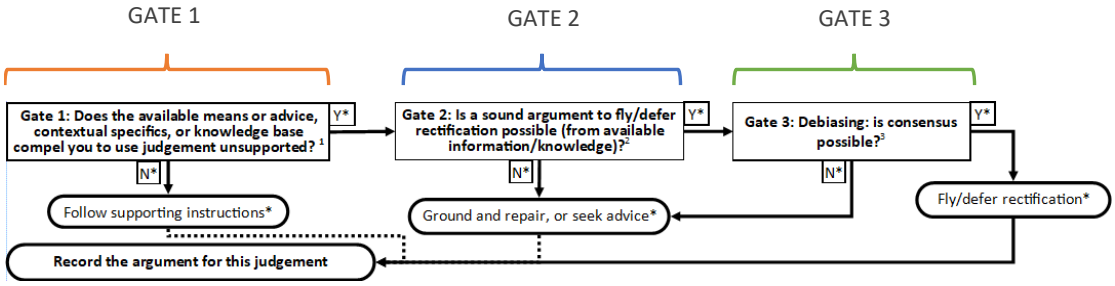
USER GUIDE AND QUICK REFERENCE GUIDE V1.0

Judgement-based Decision-making Support Framework – User guide.

Individuals make judgement-based decisions. But organisations must be accountable for their decision-making, particularly concerning risks where the potential consequences of a hazard are desired to be controlled. Airworthiness issues, structural hazards in particular, are contextually complex and decision problems are addressed accounting for a multitude of aspects, outside of the purely engineering fault consideration – for example aircraft demand, spares availability, trained manpower available. Where decision-makers are compelled to make choices as a result of time-constraints, they sometimes rely upon their judgement if there is a lack of information or resource. Judgements can be fallacious, misled by lapses in memory, making mistakes in the application of knowledge or rules to a situation, or by making a slip in the execution of their chosen action. A novel "heuristic" tree guides decision-makers in the application of their judgement toward a rational action in three distinct ways:

- Whether judgement is necessary to be applied in the current situation.
- Whether the information available to the decision-maker that is being acted upon can be structured into a logical argument to support a recommendation to fly.
- And whether a third party, who acts as an independent assessor of the argument to reduce the influence of biases, is able to reach consensus with the argument that has been formed.

If any of these answer with "no", then the decision-maker should not continue with the judgement-based decision to fly as it is more likely to contain an unchecked fallacy/error, which may lead to incident or accident, and post-hoc review may find the decision to have been made irrationally.



Gate 1: "Does the available means or advice, contextual specifics, or knowledge base compel you to use judgement unsupported?" Given the situation you are in, with the tools, manpower and resources, information, access to advice, time demand etc., are you compelled to use your judgement unsupported because you are lacking the ability to gain an analytical/authoritative solution to help with the decision-making? Unsupported means without the support of an analytical/deterministic/authoritative assessment of the issue at hand. If not compelled, then there exists authoritative guidance that should be followed.

Gate 2: "Is a sound argument to fly/defer rectification possible (from the available information/knowledge)?" Is the information/knowledge available to you able to be arranged into a sound argument that is acceptable in its form, relevant in its application and constituency, and containing good grounds that are sufficient to support a claim (in this case that an aircraft is safe to fly)? This step is covered in more detail below. If a sound argument cannot be formed from the information and knowledge available, the aircraft should be grounded, repaired or further information sought from an advisor.

Gate 3: "Debiasing: is consensus possible?" Discuss your sound argument with a peer of suitable qualification and experience; is there consensus on the argument that you've constructed? For this, the peer may use the same test as at Gate 2: whether the argument is acceptable in it's form; relevant in it's application and constituency; and containing good grounds that are sufficient to support a claim (in this case that an aircraft is safe to fly). The difference being that your peer has their own perspectives that may offer alternative insights that may strengthen or challenge the argument. If consensus cannot be gained for an argument to fly, then there is reason to doubt the soundness of the argument and the aircraft should not fly.

Gate 2 guidance – formation of a sound argument

An argument, in the classical sense, is an arrangement of information that gives logical appeal to a conclusion/claim. In it's basic form it has at least one proposition supporting another proposition. Justification by argument (in this particular format adapted from the Toulmin format) contains 6 discrete elements.

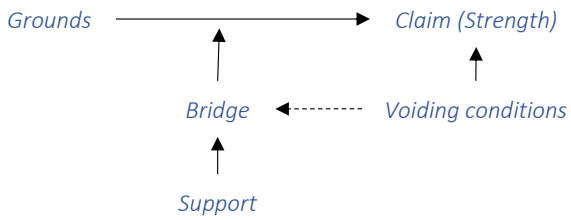
[GROUNDS]: the visible, evidential or actionable information that the decision-maker is choosing to act upon.

[BRIDGE], [SUPPORT], [VOIDING CONDITIONS]: conditional phrases that enable the claim to be reached from the *grounds for a given situation*. Without these elements, the claim may seem unreachable, logically impossible, irrational, fanciful or arbitrary. They are, for lay sake, the "requirements" or proviso of the argument. Without these elements, the argument is an undefined statement and is indefensible without resorting to stipulating information or knowledge that would form a bridge, support or voiding condition.

[CLAIM], [STRENGTH]: the valid and defensible conclusion reached from the information and knowledge available. A scale of strength of argument provides an indication for how strong the argument is and is a constructed attribute scale with the following levels:

- "Strong"**:** a link between the clauses (*[GROUNDS]*, *[BRIDGE]* and *[SUPPORT]*, *[CLAIM]*) is known to exist and is applicable to the situation being considered. There is a relevant causal chain or correlation that is accepted by consensus or authority.
- "Some"**:** the link is not strong, i.e. a link between the clauses is possible, but correlation or causality is not accepted by consensus or authority, or may be irrelevant to this situation.
- "No"**:** there is no clear link between the clauses able to be formed from available information, knowledge or data.

Visually, the interaction of the elements can be described as follows:



Gate 2 Process

Considering the function of typical information available in an aviation maintenance domain, the following process describes how the argument might be formed.

1. Recognise the extents/boundaries of the argument in regards to the *[CLAIM]* and the *[GROUNDS]*:
    a. Consider the **requirement for the claim** of airworthiness, which frames the construction of the argument:
        i. Properties of the next flight(s)
        ii. Extent of argument (for example consider the next flight, or next set of flights?)
    b. Search **information available** within decision horizon for the useful clues.
2. Select the most important information that stimulates opinion/action/deliberation on the airworthiness of the aircraft and annotate these as *[GROUNDS]*. The *[GROUNDS]* are states of nature that are observable.

EXAMPLE [GROUNDS]
a. Documentation
b. Observable fault dimensions
c. Known history of fault (on subject aircraft and fleet)
d. Known aircraft configuration
e. Known design data (including incomplete data)
f. Known aircraft usage

3. Form a statement, the *[BRIDGE]*, that follows directly from the *[GROUNDS]* and enables your valid argument about the safety of the aircraft, either as safe to fly or otherwise. With the given *[GROUNDS]*, how can a claim of safety be made? This statement has more flexibility in it's generation than the *[GROUNDS]*, as it may be formed from memory, experience about a matter, or advice provided to you. It may also contain multiple related aspects that together enable a link between the *[GROUNDS]* and the *[CLAIM]*.

EXAMPLE [BRIDGE]
a. Aircraft usage (single aircraft or fleet)
b. Maintenance history (known, inferred, advised about, recalled)
c. A mitigation option/strategy that make conditions for the claim viable
d. Design data (if actionable and perceived as relevant to the issue)
e. Individual knowledge awareness (known as meta-cognition) – knowledge of the situation/aspects involved
f. Consequence visualisation.

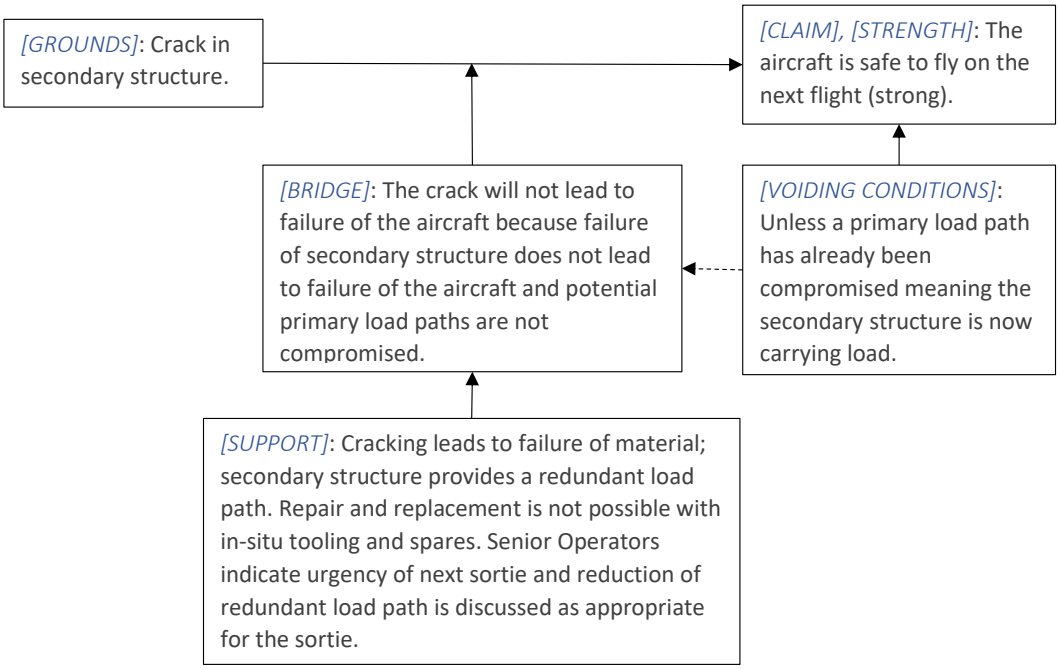
4. Provide the generalisation/categoric information that *[SUPPORT]* the *[BRIDGE]* (and only the *[BRIDGE]*). This should be formed by answering the question "why should I trust/believe/judge/accept the *[BRIDGE]* in this situation?".

   EXAMPLE [SUPPORT]

   a. Justification/necessity for mitigation (may be framed by resources available/context).
   b. Trustworthiness of information source (design data, advisor, textual information etc.).
   c. Relevance of information (to the situation).
   d. General principles, regulatory guidance or corporate accepted practices.

5. List the *[VOIDING CONDITIONS]* that would invalidate the constructed argument. These identify the conditions such that if they were true, would cause any of the bridging statements/supports to be invalid, and therefore lead to the claim also being invalidated.
6. Formally detail the *[CLAIM]* and rate the *[STRENGTH]* of the *[BRIDGE]* in relation to the *[CLAIM]*.
7. Consider the soundness of the argument – is it *acceptable* in it's form? *Relevant* in it's application and constituency, and containing *good grounds* that are sufficient to support a claim. The italicised words can helpfully be summarised as A – R – G; think of "ARGument".

Descriptive example



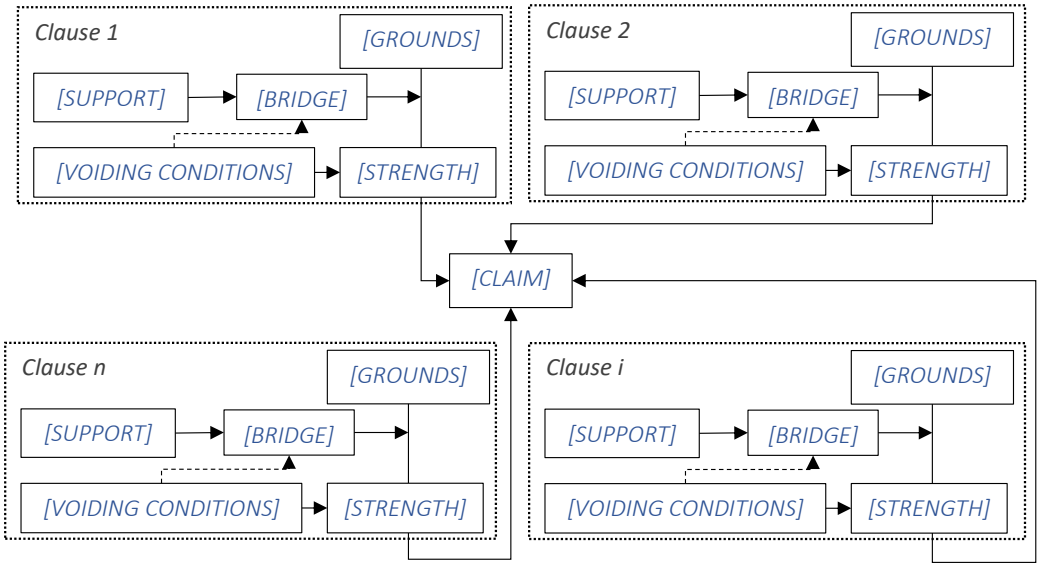
Although a generalisation about cracking may note that "cracking leads to failure of material", contextual specifics identify that the structure is classified as secondary. The risk and subsequent claim of safety is conditional on both cracking existing and the component classification. In this example, the grounds are formed of two bits of information: there being the existence of a *crack*, and that the structure involved is classified as *secondary structure*. Each of these information bits could be expressed individually to provide greater precision of communication, although this may be considered as not necessary by the decision-maker:

[GROUNDS]: a crack has been found *[BRIDGE]*: cracking means the structure is weaker. *[SUPPORT]*: cracking leads to material failure at lower than designed loads.	*[CLAIM]*: the aircraft is safe to fly the next sortie. *[STRENGTH]*: Strong. *[VOID]*: unless primary load path compromised.
[GROUNDS]: the structure involved is classified as secondary structure. *[BRIDGE]*: failure of secondary will not lead to failure of the aircraft because the primary load path has no indications of compromise. *[SUPPORT]*: secondary structure is classified by designers as providing a redundant load path.	

Equally, the decision-maker may include additional *[GROUNDS]* or *[BRIDGE]* statements in order to further satisfy a claim of safety.

[GROUNDS]: technical publications do not permit cracking on this component. *[BRIDGE]*: limits to this component are as a "catch-all" limit and generic, meaning any damage triggers OEM involvement on an "as they emerge" basis. *[SUPPORT]*: OEM are unable to provide specific limits for every type of crack or damage that may occur on an aircraft.	*[CLAIM]*: the aircraft is safe to fly the next sortie. *[STRENGTH]*: Some. *[VOID]*: unless generic limits are critically relevant to this component.

Visually represented, *n* clauses may be considered as contributing to a single judgement-based decision, represented as the argument *[CLAIM]*.



Preconception

Upon reviewing a situation, one may be intuitively drawn toward a particular claim (that the aircraft is safe or unsafe). The framework can help to guide the description of why this is so, but it can also help overturn the original perception if the argument can not be found to be sound once the *[BRIDGE]* and *[SUPPORT]* are considered more precisely. As the individual elements are populated with available information and knowledge, any contrary claims of safety will become apparent through a lack of soundness in a particular *[BRIDGE]*, *[SUPPORT]* or *[VOIDING CONDITION]*.

Risk attitudes and conflicts of judgement

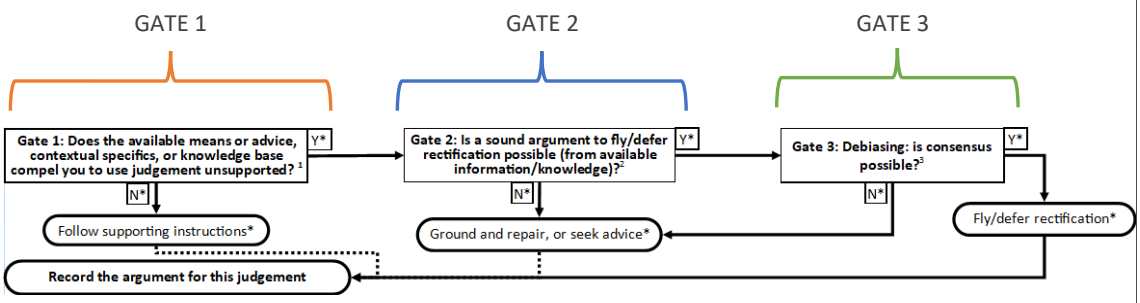
Two separate decision-makers given the same situational information may form different arguments – an alternate decision-maker to the above example may be unhappy with the reduced load-path redundancy and *[CLAIM]* the aircraft as unsafe.

If such a conflict of perspective occurs during Gate 3, the argument should be further decomposed (if possible within the situational constraints), identifying additional grounds, bridges, supports or voiding conditions to satisfy the need to validate an argument. Else, consensus can not be established and the aircraft should not be flown with the presented judgement-based argument. Applying argument to judgement must ultimately encourage sufficient decomposition of the judgement to enable appropriate communication, audit trail and satisfactory justification given the situational constraints.

Arguing ground/repair/seek information positions

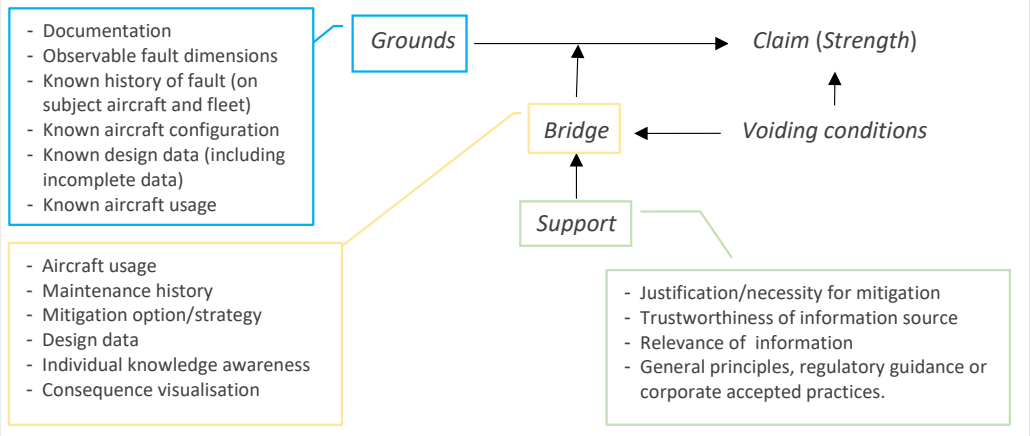
If a user answers “N” to any of the Gates, they may wish to justify their argument using the format provided. Depending upon communication requirements, it may be requisite to substantiate any claim regarding airworthiness.

QUICK REFERENCE GUIDE



GATE 1: does the situation compel you to use your judgement? If not, then follow the supporting documentation/instructions already available.

GATE 2: Formation of a sound argument given the information and knowledge available in the situation.



Gate 2 Process:

1. Recognise the extent of the argument. limit
2. Select key information for *[GROUNDS].*
3. Construct the *[BRIDGE]* that follows from the *[GROUNDS]*.
4. Detail the *[SUPPORT]* for the *[BRIDGE]*.
5. Identify *[VOIDING CONDITIONS]* that invalidate the stated argument.
6. Express the *[CLAIM]* and *[STRENGTH]*.
7. A-R-G; review the soundness of the argument.

GATE 3: Use a peer/colleague to review your judgements and seek consensus. If no consensus is possible, and the argument cannot be adjusted to account for additional perspectives in support of a claim to fly, then the aircraft should be grounded until the argument can be improved with further information or advice, or the aircraft is repaired.

Notes supporting Judgement-based Decision-Making Support Tool

Note 1. Using judgement "unsupported" means that there is/are no authoritative source(s) on which to base a decision regarding an identified fault.

Note 2. A *sound* argument can be formed when the information available to the decision-maker in a situation can be validly organised to support the claim or conclusion. A sound argument is formed when information *[GROUNDS]* is linked to a claim by a bridging licence *[BRIDGE]*, and supported by categorical or generalisable information/knowledge *[SUPPORT]*. The licences used in an argument may be context specific. A *strong* argument can be formed when a sound argument is supported by a statement of the conditions under which the argument becomes void *[VOIDING CONDITIONS]*, and by an indication of the strength of the argument *[STRENGTH]*. Arguments can be found to be sound by reviewing their:

· *Acceptability*: it is acceptable to those receiving the argument (A).

· *Relevance*: the premises are relevant to the conclusion (R) and all relevant information has been taken into consideration.

· *Good grounds*: It has sufficient or good grounds for the conclusion (G).

Note 3. Individualistically formed judgement can be biased and difficult to debias without support from experts or peers of similar repute within the domain that the argument is formed around. Consensus must be formed around the formed argument with suitably reputable peers.