EngineeringX

# Safer Complex Systems
## An Initial Framework

Royal Academy of Engineering

Lloyd's Register Foundation

# Safer Complex Systems

## An initial framework for understanding and improving the safety of complex, interconnected systems in a rapidly changing and uncertain world

J. A. McDermid OBE FREng
S. Burton
P. Garnett
R. Weaver

University of York
Heslington, York, UK

Royal Academy of Engineering
Prince Philip House
3 Carlton House Terrace
London SW1Y 5DG

Tel: 020 7766 0600
www.raeng.org.uk
@RAEngNews

ISBN 978-1-909327-49-8

UNIVERSITY of York

ASSURING AUTONOMY
INTERNATIONAL PROGRAMME

# Contents

*The views and opinions expressed in this report are those of the University of York research team and do not necessarily reflect the views of Engineering X.*

# 0

# Foreword

Our lives, society, industry and commerce all depend on a range of systems that are becoming ever more complex, interconnected and interdependent. The growth in complexity of designed systems and the emergent complexity of ad hoc systems are outstripping our engineering methods and challenging our ability to manage systems safely. A range of factors is bringing us to a tipping point. Urgent action is needed to ensure that the levels of safety society has come to expect are preserved into the future.

# Foreword by the Royal Academy of Engineering

**Dame Judith Hackitt DBE FREng**
**Chair of the Engineering X Safer Complex Systems mission**

The world is a different place from when we started the programme. In early 2020 the COVID-19 pandemic shifted the frame of reference for all of us and demonstrated how the increasing complexity and interconnectedness of the world we live in has made us all more vulnerable to systemic shocks of this nature. A major global safety challenge, therefore, is to develop our understanding of the root causes of systemic failure and to take collective action to prevent or mitigate against future events with a similar potential to harm.

The Safer Complex Systems mission aims to increase the safety of complex infrastructure systems globally. We began with an initial workshop in July 2019 where we brought together 40 experts from different sectors and disciplines to share thoughts on complex system safety. Workshop participants confirmed our views that the safety management tools of the past were no longer appropriate for the complexity of the issues we currently face. They also suggested we commission a piece of research to help understand what we already knew about safety and complex systems, what the existing tools available to us in different sectors are and how knowledge is currently shared between them.

In December 2019 we commissioned the University of York team to undertake this research. I want to thank the authors for their considerable work during the most testing of times and the members of the Technical Advisory Group, who diligently and enthusiastically fed back on earlier drafts. I also thank Lloyd's Register Foundation, our funding partner for this programme and co-founder of Engineering X.

This report is an important first step in the journey to helping society to better manage complexity. Through powerful cases studies, it highlights the need to systematically understand all of the components, their interconnectedness, and the non-linearity whereby a simple failure in one place can lead to a surprising and disproportionate effect elsewhere.

Importantly, the report also makes a start on creating a framework to enable people in different sectors and disciplines to talk about, analyse, and manage safety in different contexts. We have already tested the framework in two global workshops held virtually in September 2020, with 131 participants from 20 countries across six continents. We were very pleased that attendees found the framework broadly helpful and affirmed again the value of our endeavour. But it was also clear that there's much more to be done! For example, testing this initial framework to make it as useful as possible in different global settings and using the findings to create educational products and practices that assist safety managers to do their job in a changing environment.

Our experience of the COVID-19 pandemic tells us that this work to enhance safety in complex systems is more important and urgent than ever. As we begin to rebuild in a post-COVID-19 world we need to ask ourselves three questions: how can we manage complexity more effectively? How can we find way to simplify and share knowledge? And how do we raise awareness and competence? There is much more we need to do to increase competency in systems thinking and share good practice across engineering disciplines and beyond. We hope that this report helps you think further about these issues and that we can work together to engineer safer outcomes in an increasingly complex world. Join us.

# Foreword by Lloyd's Register Foundation

**Dr Jan Przydatek**
**Director of Technologies, Lloyd's Register Foundation, and Board member, Engineering X Safer Complex Systems mission**

When we published our insight work on global safety challenges, the findings on safety of complex interconnected systems added to growing evidence highlighting the need for practical interventions that would enable the continued safe operation of these systems as complexity increases.

The world we live in is made up of many interconnected systems that, in combination, create the critical infrastructures and supply chains that we all rely upon. Each individual system plays its part, influencing the other systems that are directly or indirectly connected to it. When all the systems work together as a complex system, they deliver an outcome that is often a service, from which we benefit.

When we have a good understanding of how these complex systems work, we can control them and be confident that they will deliver the services we expect. But complex systems can grow and evolve over time and our inability to understand them means that unexpected behaviours can develop; unanticipated outcomes can ultimately threaten life and property.

This report was commissioned to inform the Safer Complex Systems programme and the wider community on how it is possible to enhance the safety of these systems. It introduces a new framework to manage the risks that are associated with complex systems (with some industry-specific examples of the method that can be widely applied across other sectors and geographies) and recommends key themes where the programme could focus to make a distinctive difference.

Lloyd's Register Foundation is an independent global charity with a mission to engineer a safer world. We know that global challenges need global solutions; they cannot be tackled by working alone. So, together with the Royal Academy of Engineering, we founded Engineering X to bring together a global community to tackle some of the greatest challenges of our age.

# Acknowledgements

The authors would like to express their sincere gratitude to all those who contributed to this study. This includes the workshop participants, questionnaire respondents and the Technical Advisory Group (TAG) who provided extremely valuable feedback on earlier drafts of this report.

Also, particular thanks go to colleagues at the University of York who assisted in the study, specifically Manogna Goparaju and Yan Jia who worked on the supply networks and healthcare aspects of the study, respectively.

Engineering X and the report authors would like to express their sincere gratitude to all those who contributed to this study, including:

## University of York

**Professor John McDermid OBE FREng**
**Director of Assuring Autonomy International Programme**
University of York

**Dr Philip Garnett**
**Senior Lecturer in Systems and Organisation**
University of York

**Professor Simon Burton**
**Honorary Visiting Professor**
University of York
**Research Division Director**
Fraunhofer Institute for Cognitive Systems

**Dr Rob Weaver**
**Global Aviation, Safety and Risk Advisor**
University of York

## Technical Advisory Group

**Prof Roger Kemp MBE FREng (TAG Chair)**
**Emeritus Professor**
Lancaster University

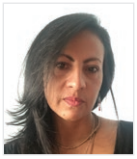
**Professor Robin Bloomfield FREng**
**Professor of Software and System Dependability**
Adelard LLP and City, University of London

**Professor Helen Atkinson CBE FREng**
**Pro-Vice-Chancellor of the School of Aerospace, Transport Systems and Manufacturing**
Cranfield University

**Professor Nilay Shah FREng**
**Head of Chemical Engineering**
Imperial College London

**Dr Joanna Boehnert**
**Lecturer in Communication Design**
Loughborough University

**Dr Tingting Zhu**
**Royal Academy of Engineering Research Fellow**
Oxford University



**Diana Torres**
**Director**
Suava Fundación

**Dr Dougal Goodman OBE FREng**
**Vice President**
The Foundation for Science and Technology

**Steve Yianni FREng**
**President**
AIRTO

**Professor Philip John FREng**
**Emeritus Professor**
Cranfield University

**Matt Crossman**
**Team Leader**
National Infrastructure Commission

**Dr Chris Elliott MBE FREng**
**Director**
Pitchill Consulting

**Shahana Buchanan**
**Global Head of Environment, Safety and Health**
Sulzer

**Professor Seth Blumsack**
**Professor of Energy Policy, Economics and International Affairs**
The Pennsylvania State University and Santa Fe Institute

**Professor Diane Finegood**
**Professor and Fellow**
Morris J. Wosk Centre for Dialogue, Simon Fraser University

**Michael Macdonnell**
**Director of Global Deployment**
Google Health

**Dr Chris White**
**Senior Programme Manager**
Lloyd's Register Foundation

**Professor Chan Ghee Koh**
**Director of Lloyd's Register Foundation Institute for the Public Understanding of Risk**
National University of Singapore

**Professor Richard Parker CBE FREng**
**Chairman**
Singapore Aerospace Programme

**Kerry Lunney**
**Country Engineering Director**
Thales Australia
**President**
International Council on Systems Engineering (INCOSE)

**Dr Guru Madhavan**
**Norman R. Augustine Senior Scholar and Senior Director of Programs**
US National Academy of Engineering

## Technical Advisory Group (continued)

**Benjamin Kumpf**
**Head of Development Innovation Team**
OECD

**Professor Jonathan Dawes**
**Professor of Applied Mathematics**
University of Bath

**Professor Gary Burnett**
**Professor of Transport Human Factors**
University of Nottingham

**Professor Brian Collins CB FREng**
**Professor of Engineering Policy**
University College London

**Dr Rainer Groh**
**Royal Academy of Engineering Research Fellow**
University of Bristol

**Professor John Clarkson FREng**
**Director of the Cambridge Engineering Design Centre**
University of Cambridge

**Professor Liz Varga**
**Professor of Complex Systems**
University College London

**Professor C.C. Chan FREng**
**Honorary Professor and the former Head of the Department of Electrical and Electronic Engineering**
University of Hong Kong

## Engineering X

**Shelley Stromdale**
**Programme Manager, Safer Complex Systems Engineering X**
Royal Academy of Engineering

**Shaarad Sharma**
**Senior Manager Engineering X**
Royal Academy of Engineering

## Design and Photography

**Joanna Boehnert**
**Design direction and figures**

**Orlagh O'Brien**
**Graphic and cover design**

**Peter Skyoto, SystemViz**
**Icons** Figure 1

**Joshua Fuller**
**Photo** Aerial view of Dubai

**Belkos**
**Photo** Flying aircraft

**Chuttersnap**
**Photo** Container Port in Bukit Merah, Singapore

**Halfpoint**
**Photo** Patient in hospital bed

# Executive summary

Our lives, society, industry and commerce all depend on a range of systems, including critical infrastructure, that are becoming ever more complex, interconnected and interdependent. This report, produced through the Engineering X Safer Complex Systems programme, has three primary objectives. **First, to develop conceptual clarity around what is meant by Safer Complex Systems and to produce a framework that provides a common basis for discussing the safety of such systems. Second, to assess the effectiveness of the existing methods available for the design, management and governance of complex systems. Third, to outline emerging challenges and opportunities with significant disruptive potential (negative or positive) for the safety of complex systems.**

The report draws on complex systems theory and real-world experience of complex systems engineering and operation to provide an initial principled but pragmatic framework for complex systems safety management. The framework spans governance (including regulation), management (of individual systems) and task and technical layers. It identifies causes of complexity, their consequences and then provides an initial classification of the systemic failures that can arise from unmanaged complexity. Further, the framework identifies design-time and operation-time controls that have a role in managing safety. The framework was developed following extensive stakeholder engagement and analysis of accidents and incidents across several domains.

In some domains, effective long-term management and governance of safety has led to sustained reduction in accident and fatality rates. Therefore some classes of system are remarkably safe, but this finding is tempered by the impact that some industries have on the environment and consequently on human health. Additionally, systems are growing in complexity and are becoming more widely interconnected and interdependent. The interdependency of apparently independent systems has been starkly exposed by COVID-19. **The growth in complexity of designed systems and the emergent complexity of ad hoc systems are outstripping our engineering methods and challenging our ability to manage systems safely.** A range of factors are bringing us to a tipping point, such as modern technologies including communications, artificial intelligence and autonomy, and commercial practices including globalisation and casualisation of labour. Urgent action is needed to ensure that the levels of safety society has come to expect are preserved into the future. There is a need to develop new methods of risk management and to integrate them with existing successful methods where they remain relevant. There are several key threads to achieving safety including: resilience; agility in safety management; safety culture, with a focus on equality, diversity and inclusion; and the use of data both to understand accident causation and to identify leading indicators of problems to help in their management.

The recommendations of the report are primarily focused on the Safer Complex Systems programme itself as follows:

- **Develop approaches for better communicating risk, increasing trust and forming consensus on acceptable levels of safety**
- **Acknowledge and address complexity in oversight, regulatory structures, legal accountability and policymaking**
- **Develop methods to address equality, diversity and inclusion during risk management and promote heterogeneity of thought**
- **Integrate simulation, model-based analysis and digital twins into design and operational-time controls**
- **Develop an integrated and complementary set of methods for analysing risks in complex systems**
- **Identify design-time and operation-time controls for increasing system resilience**

# 1

# Introduction

This section introduces the aims and objectives of this study, which will inform the future direction of the **Safer Complex Systems** programme. It presents the three layers (governance, management, task and technical) that form the framework developed in the study.

We rely on critical infrastructure and sociotechnical systems to keep us healthy and for the economy to flourish, increasing standards of living across the global community. These systems, which our daily lives depend on, are becoming ever more complex and interconnected. The failure of such systems to meet key objectives can have both direct and indirect consequences for the safety of people and the environment as well as for trust and public confidence in the systems. The COVID-19 outbreak has demonstrated the challenges that society faces in a globally connected world. This pandemic has not only had a direct impact on the health of millions of people, but also indirectly impacts society in an unprecedented manner by disrupting the complex systems that support almost every facet of daily life from supply chains, food production, transport and education to social interactions and the economy as a whole. This study, commissioned by **Engineering X**, a new international collaboration founded by the Royal Academy of Engineering and Lloyd's Register Foundation, has been tasked with examining which measures are required to reduce risk and increase safety by improving the design, management and governance of complex systems. The objectives of the study are to:

1. develop conceptual clarity around what is meant by **Safer Complex Systems** by producing a framework that provides a common way to communicate about the safety of complex systems across sectors and between different levels of expertise globally
2. develop understanding of the existing methods available for the design, management and governance of complex systems (including those developed in academia that have not yet been implemented)
3. outline emerging challenges and opportunities with significant disruptive potential (negative or positive) with regards to the safety of complex systems.

Ensuring the safety of increasingly complex systems is challenging. In particular, unacceptable levels of risk will occur if the complexity of the systems and their operating environments outpace our ability to engineer, operate and govern such systems – urgent action is needed to prevent this from happening. It must also be a guiding principle to make the world safer for everyone regardless of age, gender, race, ability, sexual orientation, religion, belief, nationality or social status. This study also considers the impact of equality, diversity and inclusion in managing the safety of complex systems. In addition, the study also addresses heterogeneity as a key tool in managing risk by ensuring that diverse perspectives and skills are considered when designing and operating these systems.

System complexity inevitably leads to gaps between intended or desirable properties and what is specified and subsequently implemented. Such gaps should be considered not only from technical but also from legal and ethical perspectives [1]. The framework developed in this study identifies measures and controls to address and reduce these gaps and to manage risk – in many cases the levers are regulatory or financial, instead of, or in addition to, technical. However, not all systems are explicitly engineered; they can also ensue through ad hoc interactions between components previously considered unrelated. This requires radically different approaches and viewpoints to previously applied safety engineering and management techniques that were based on clearly defined system boundaries. The framework presented in this report is intended to address both designed and *ad hoc* systems and is structured into the following layers:

- **Governance** – This layer consists of incentives and requirements for organisations to adhere to best practice through direct regulation, so-called soft law approaches or a consensus in the form of national and international standards. In formulating these standards and regulations, governments and authorities represent societal expectations on the acceptable level of residual risk that is to be associated with the systems.
- **Management** – This layer coordinates tasks involved in the design, operation and maintenance of the systems, enabling risk management and informed design trade-offs across corporate boundaries, control over intellectual property and liability, management of supply chain dynamics and sustaining long-term institutional knowledge for long-lived and evolving systems.
- **Task and technical** – This layer covers the technical design and safety analysis process that allows systems to be deployed at an acceptable level of risk, then actively monitored to ensure deviations between what was predicted and what is actually happening so that these gaps can be identified and rectified. This layer includes not only the technological components but also the tasks performed by the users, operators and stakeholders within a sociotechnical context. In some cases, users may be unwilling or unknowing participants in the system who are nevertheless impacted by risk.

Interactions between these layers can be one of the major causes of system complexity, leading to systemic failures that may not have been predictable if considering a

single layer and that may not be controllable at a single layer.

This study is the result of preliminary work in this area, so will not define a complete and validated methodology for ensuring safer complex systems. The scope of the work has been deliberately limited, specifically to a small number of application domains and with a focus on safety. However, the report seeks to provide structure and direction for further work in this area to both mature the framework and expand the scope of its applicability.

Section 2 begins by providing a number of definitions used to focus the study. The framework itself is presented in detail in Section 3, including initial ideas for a set of control measures to manage safety risk – recognising that current methods are not sufficient in the face of growing complexity. These measures can be divided into those that are (most) relevant at design-time versus those that are relevant at operation-time. Section 4 analyses the state-of-the-art in safety management and analysis techniques to gauge their applicability to explicitly addressing causes and consequences of complexity. Section 5 summarises the industrial sectors that were considered to inform this phase of the work. Section 6 presents the key findings from the study based on the analysis in the preceding sections and the results of stakeholder engagement.

The report concludes with a set of recommendations for future work, including how the framework presented could be applied to further sectors with Engineering X selecting case studies for further investigation.

In performing this study, the authors made use of considerable stakeholder engagements, academic research and their own experience in a wide range of industrial sectors. Appendix B summarises the results of the stakeholder engagements and Appendix C extracts insights from cases studies from the aerospace, mobility, healthcare and supply chain sectors, which were used in developing the framework. Some sector-specific recommendations are set out in Appendix D.

The framework presented addresses both designed and *ad hoc* systems and is structured into three layers: governance; management; and task and technical. Interactions between these layers can be one of the major causes of system complexity.

# 2

# What is a safe complex system?

In this section we define some important concepts that are required to help structure and scope the study. A key objective of the study is to develop conceptual clarity around what is meant by **Safer Complex Systems** by producing a framework to support a common way to communicate about the safety of complex systems across sectors and between different levels of expertise globally. Our approach has been to draw on existing concepts and definitions, so far as is practicable, but to adapt them to draw together a coherent set of concepts. In the interest of brevity, not all concepts and terms are defined in this section. A comprehensive set of definitions of the concepts used in this report can be found in Appendix A.

# 2.1 Definition of a complex system

In this study we define a *system* as an arrangement of parts (or elements) that together exhibit behaviour or meaning that the individual constituents do not. Complex systems theory would define a system as *complex* if some of the behaviours of the system are emergent properties of interactions between the parts of the system, where behaviours are unpredictable from knowledge of the parts and their interactions alone. This is often reduced to the phrase "the whole is greater than the sum of the parts" [2].

Complexity is therefore not determined by the size of the system or the number of parts and interactions, but is instead determined by the nature of the interactions between the parts and the relationship between the parts and their environment. Relatively small systems with few parts and possible interactions can exhibit emergent properties of the interactions between the parts of the system and are therefore, by definition, complex systems. A good example would be Conway's Game of Life [3]. Small systems using machine learning, especially where the learning continues into operation, are also likely to be complex. However, this study will also consider complex systems at a larger scale, including those that illustrate the interaction between society and technology such as healthcare services, supply networks and transport.

From the perspective of complexity science there are a few characteristics that are shared by most, if not all, complex systems. These are variously described and defined in [4, 2, 5]. The primary characteristics of interest in this study are summarised in Figure 1. Definitions of these characteristics can be found in Appendix A.2, and two key characteristics are illustrated in some detail here.

The boundaries between the system and the environment are dependent on the scope of the system under consideration, known as the *system of interest*. This may vary depending on the objectives of the analysis. For example, if focusing on the functional performance of an automated vehicle the system could be viewed as a set of electronic components, which sense the environment, decide on control actions and implement them via actuators. However, when considering a mobility service, the system would include other traffic participants and city infrastructure as well as the vehicle. In short, it can be challenging to determine where the system ends and the environment begins, and that boundary may change over time. This also presents a challenge for those responsible for complex systems, as they may find that the system that they have some duty or responsibility to manage crosses over multiple *boundaries* (for example, jurisdictions, borders, domains, or organisations).

Defining the boundary of a complex system requires an explicit decision about what is inside the system and what is out of scope. Such decisions are inevitably performed within a social context and are therefore subject to the bias of those defining the system (the decisions can be seen as political [6]). Complex systems therefore cannot be considered without considering their context, and different stakeholders may perceive the 'same' system in different ways.

System boundaries are therefore frequently defined pragmatically, drawn to be representative of the *system of interest* and mark the limits of that system. It would normally be possible to draw a different boundary, with different limits, which could depend on the society in which those drawing the boundary originate. Boundaries are also permeable: there are numerous feedback paths between the environment and the *system of interest*.

There is also the possibility that *systems* can emerge in an ad hoc way, through a convergence of parts perhaps by a process of self-organisation, or self-assembly. Here the semi-permeable boundary may present itself as an emergent property of that system, and change as the system evolves.

It is common to talk about complex systems going through critical transitions [7] widely referred to as *tipping points*, Figure 1. These are a form of rapid transition from one system state to another [8], possibly from a stable state to one of instability, or a change in purpose. Tipping points can be brought on by shocks to the system. The response of many organisations to the COVID-19 pandemic are examples of tipping points. Tipping points can be positive, as they represent the ability of a system to respond successfully to environmental change. However, tipping points can also be transitions into an unsafe state, and these can be emergent properties of the system itself. For example it is often hard to precisely determine why a crowd of protesters transitions into a riot, but it could be an emergent property of the internal dynamics of the crowd or an external shock.

As well as tipping points, the temporality of risk in complex systems can also emerge over time as complex systems interact with their environment. For example, the damaging effects (both to the environment and health) of the pollution caused by the automotive industry does not manifest in a single car at a single point in time, but instead is a consequence of populations of cars and how and where they are used through time. These longer-term consequences of complex systems are difficult to identify until they manifest in the environment. An example of long term environmental impacts of complex systems, focusing on PFAS chemicals, is discussed in section C.4.2.

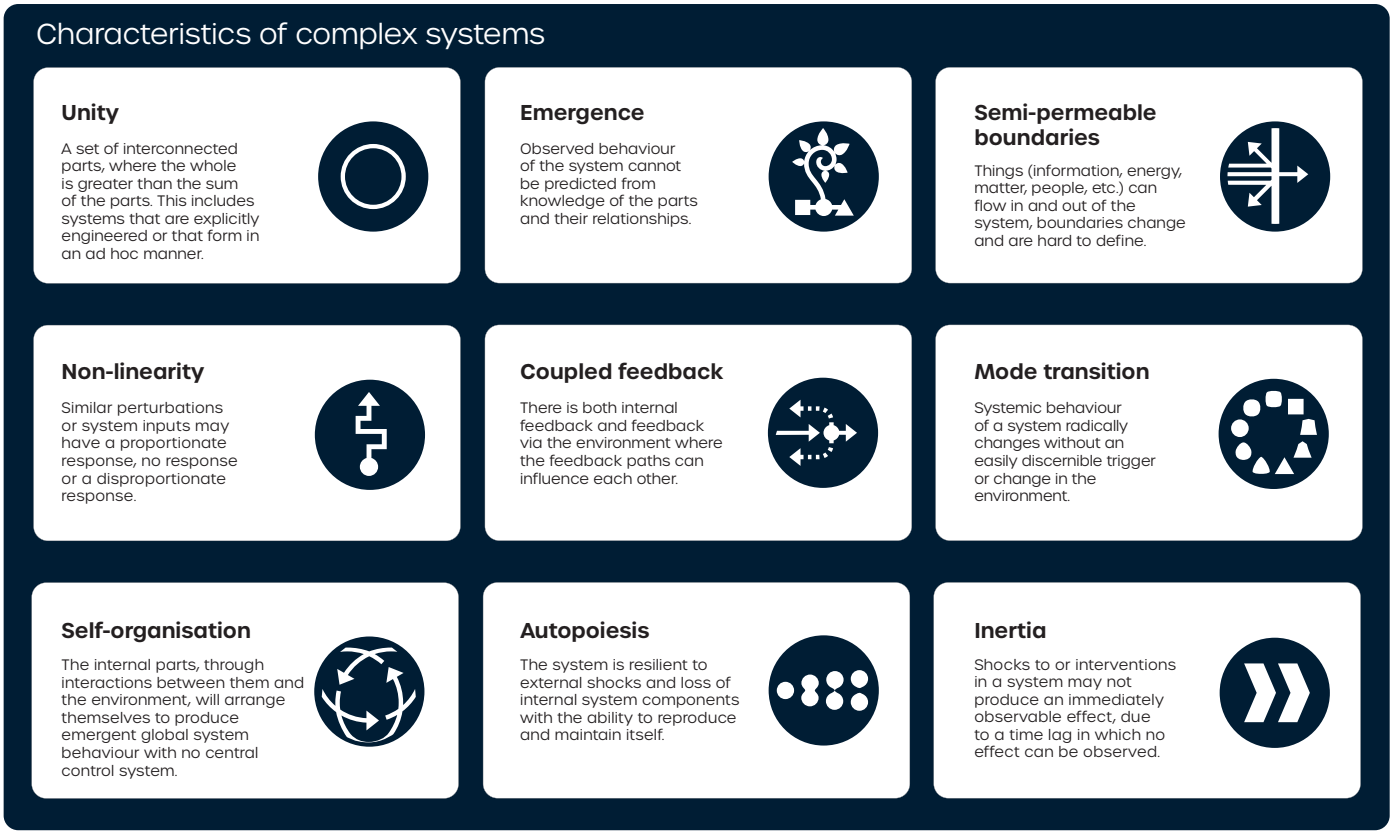



Figure 1: Characteristics of complex systems

The features of complex systems such as *non-linearity*, *self-organisation* and *emergence* can make it difficult to predict that a system is close to a tipping point, or to determine what triggered the tipping point. However early warning signals have been identified for some systems that give an indication that the system could be nearing a tipping point [9, 10, 7]. Determining what comes after a tipping point is even more difficult and some might say impossible [11].

These features of complex systems make them difficult to precisely control as it is hard to develop a prior understanding of how a system is going to react to an intervention, and there is no guarantee that a system will behave in the same way in reaction to multiple similar interventions. It is often therefore only through observation (and to an extent, simulation) that we can learn how a system is going to react to our attempt to govern it. This can at best be a cause of frustration, as the normal levers of management could produce little or no discernible effect. Worse, an intervention intended to drive the system to a safe state could result in the opposite happening – perhaps the Boeing 737 MAX accidents are a good illustration of this, see Appendix C.1.2. Regardless of whether the outcome is good or bad, unexpected behaviour is not conducive to safety.

Various metaphors have been used to describe how we should change the behaviour of complex systems through interventions. It is often described more in terms of a continuous process of stewardship, where relatively modest interventions are made to guide a system in a particular direction [12]. As systems become more complex we might therefore have to make rather modest and continuous changes to both the system (perhaps an engineered device) and its context/environment (how it is used, and how its use is governed) to steer it in a desirable direction (a process of convergence). This is likely to be much more effective than making large changes to a system and its environment and expecting the system to behave as predicted.

The process of stewardship, or the steering of systems captures some aspects of what is often referred to as systems thinking or complexity thinking. These modes of thinking seek to move away from a mechanistic, or reductionist, approach to knowledge production with simple (often linear) causal paths. Complex thinking is a mode of thinking that attends to, and integrates, some aspects of the complexity of the world [13]. How this can be productively applied to *engineered* and *safe* systems is an area in need of further work. This framework is a step towards this, as it draws attention to the causes and consequences of complexity within systems.

This framework (see Section 3) builds on these characteristics of complex systems to identify causes and consequences of complexity, as well as exacerbating factors that contribute to systemic failures and safety risks.

# 2.2 Safety, risk and systemic failure

Risk is a widely used term, but the definition varies between domains. Risks can also relate to different categories of objective, including business, mission, societal, environmental and dependability, which includes safety. This report focuses on risks related to safety. There are three related terms:

- **Safety** – freedom from unacceptable risk.
- **Risk** – the combination of the likelihood of harm and the severity of harm.
- **Harm** – physical injury or damage to the health of people.

It is important to note that harm can arise either directly, or indirectly, for example through effects on another system or via damage to property or to the environment. Further, the term "intolerable" is sometimes used rather than "unacceptable".

Complex systems failures can have many different forms, for example loss of power or communications capability, or disruption to food supply chains. However, for this to be considered a *safety* issue there has to be an impact on human health rather than, say, a purely economic loss – the safety impact can be indirect. For example, environmental impacts of emissions from road vehicles, electrical power plant or factories impact human health, hence pose an indirect safety risk. Alternatively, we could say that the *system of interest* includes the environment – but often it is more helpful to discuss the narrower systems, such as an electrical power plant, to ensure a clear focus for analysis. The risk can also be long-term or cumulative – for example asbestosis and other chronic health conditions often only arise over many years of exposure.

Realistically, no system or situation can be risk-free and it is often necessary to compare risks between different systems or situations, noting that doing nothing can also increase risk, for example in a healthcare situation or in decommissioning a nuclear power plant.

It is often hard to quantify the elements of risk, so engineering practice typically applies qualitative measures of severity. These could look like this:

- **Catastrophic** – multiple fatalities or serious injuries.
- **Critical** – single fatality or a small number of serious injuries.
- **Major** – a single serious injury or a small number of minor injuries.
- **Minor** – a single minor injury.

Here, serious injury means something that has a lasting effect, such as the loss of a limb, whereas a minor injury is normally totally recoverable.

Likelihood can also be defined qualitatively, although it is more common to seek to quantify it, for example in terms of failure probability per operational hour.

Traditional system safety engineering focuses on component failures and their interactions. In contrast, complex systems can give rise to systemic failures that are distinct from failures of individual system parts, namely a failure at the system level, rather than failure of a particular system component.

This bears a strong and deliberate relationship to the definition of complex systems and the notion of emergence. Failure can be partial – a system can keep operating but lose particular functions or capabilities. For example, the loss of the UK's Air Traffic Management (ATM) capability in December 2014 did not prevent other parts of the worldwide air transport system, including aircraft and airports, from operating, see Appendix C.1.1.

Care must be taken to distinguish between failures that are systemic versus those that are not. Systemic risks originate from the interactions between the parts of the system (their behaviours) and interaction with or dependencies on the environment, rather than poor assembly/manufacturing, faulty components, buggy software functions or wear and tear – although such things might be triggers for a systemic failure. Take, for example, an aircraft accident such as a tyre bursting on landing. If this was due to failure of parts caused by poor engineering or manufacturing, leading to the seizure of a brake so that the wheel could not rotate when the aircraft landed, then we would not class this as a systemic failure. On the other hand, the failures of the Boeing 737 MAX (see Appendix C.1.2), can be viewed as systemic in that the unsafe behaviour was an emergent property of the system, including the pilots. This type of failure might not be predicted from an understanding of the parts of the system and their failures alone, however it could potentially be discovered through a simulation involving pilots in responding to failure events.

Traditional safety engineering (see Section 4), uses categories of failure modes to help structure and guide the safety process. For example, in functional hazard assessments, it is common to consider the following modes of failure:

- **Omission** – function not provided when intended.
- **Commission** – function provided when not intended.
- **Incorrect** – malfunction, for example production of an incorrect value.

To our knowledge, no such categorisation of systemic failures exists. Section 3.5 explores the idea of categorising systemic failure and makes some initial suggestions with the expectation that these will be reviewed and refined in later phases of the **Safer Complex Systems** programme.

Acceptability of risk is a social construct and it may involve comparison of different options including doing nothing which, as noted above, can lead to increases in risk. In many domains, acceptable levels of risk are codified either as numerical targets, for example for civil aircraft, catastrophic events shall not occur more than one in every billion flying hours, or qualitatively, for example using risk matrices. There is also a range of risk acceptance principles, often enshrined in law (see Section 4.2). Although the details of these principles vary, they are all focused on ensuring the acceptability of residual risk, that is the risk of harm that remains once risk reduction measures have been implemented.

Risk reduction measures, or controls, can include engineering changes at design time, or procedures and processes implemented during operation. Controls can be grouped in very broad terms into those that enable:

- **Robustness** – the ability of a system to cope with foreseen events.
  This is contrasted with:
- **Resilience** – the ability of a system to absorb the unforeseeable.

Both resilience and robustness are tools for reducing risk, with resilience more important in dealing with the uncertainties arising from complexity. Complexity science uses these terms rather differently. For example, resilience is used to mean that the system returns to its original state. Here we deliberately use a definition that would include reducing capability while remaining safe. This is quite common in operational safety management, for example see the UK ATM failure in Appendix C.1.1, and it is common to identify safe 'fallbacks' for dealing with failure situations. The notion of harm and residual risk remain valid for complex systems but it is far from clear whether or not the established approaches to determining acceptability of risk are still appropriate. One further dimension that needs to be considered is that of 'trust', which we define in this context as firm belief in the reliability, safety or capability of a system, individual or organisation.

For complex systems this specifically means the trust that stakeholders have in the system or in the organisations responsible for it (including developers, operators and regulators). Decisions about acceptability of a system might be based on trust rather than a more direct assessment of risk. Trust can be a greater issue for complex systems where the behaviour is emergent and unexpected. Unexpected outcomes that cannot be fully predicted are likely to cause a (potentially significant) reduction in system trust and acceptance by users and other stakeholders.

# 2.3 Responsibility, accountability and safety culture

In some contexts, it is common to use the term *duty holder* to indicate that a person or role is responsible for risk (including ensuring an acceptable level of residual risk). The duty holder can be ultimately held accountable for risk caused by a system. The following definitions are derived from the Oxford English Dictionary, adapted slightly to draw clearer distinctions between the concepts.

- **Responsibility** – having a duty to ensure that actions are taken or avoided to reduce risk.
- **Accountability** – required or expected to justify actions and decisions.

Normally, to be accountable, one would have to be responsible for those actions and decisions, but accountability might only exist for senior roles in an organisation, and someone in those roles would have vicarious responsibility for the acts, decisions (and omissions) of subordinates. A person could be held accountable for the actions of others where they have responsibility and authority, and it would be reasonable to expect a duty holder to have (the relevant) authority.

- **Authority** – the power or right to give orders, make decisions, and enforce obedience.

One of the ways in which vicarious responsibility manifests itself within organisations is in the form of a safety culture. A lack of safety culture can lead to significant safety risks being introduced or overlooked due to the lack of collective responsibility. The UK advisory committee on the safety of nuclear installations defines safety culture as follows: "the **safety culture** of an organisation is the product of individual and group values, attitudes, competencies and patterns of behaviour that determine the commitment to, and the style and proficiency of, an organisation's health and safety management".

One of the consequences of system complexity can be a difficulty in assigning accountability to a single person, role or organisations. Further, it may not be appropriate to hold someone to account even if they were (formally) accountable for something but they did not have moral responsibility. Moral responsibility only occurs if two conditions are met [1]:

- **Control condition** – the person must have relevant control over the action, such that the action adequately represents or reflects the person's intentions or desires.
- **Epistemic condition** – the person must have had relevant knowledge and understanding of the action, and its likely consequences.

These conditions illustrate that the topic of *safety culture,* and in particular the *accountability* of people and organisations, must be considered when analysing complex systems. These can be thought of as design considerations for the organisations developing, operating and regulating complex systems. We anticipate that changes will be needed in the way accountability and responsibility are defined and allocated to deal with the challenges of complex systems, although details of such considerations are outside the scope of this report.

# 3

# A framework for safer complex systems

The aim of this section is to give an overview of the framework, to discuss its maturity and to indicate ways in which it might evolve. However, as there are a lot of elements in the framework, the text here focuses on the structural aspects of the framework and details are presented in Appendix A.3. This section starts by outlining some of the requirements for the framework before introducing the framework itself.

# 3.1 Requirements and response

The purpose of the framework is to provide conceptual clarity around the factors that lead to systemic failures that have a safety impact in complex systems, as well as appropriate measures for minimising or mitigating the resulting risks.

The framework is intended to provide a structure and concepts for supporting the communication of these issues across sectors and between different levels of expertise globally. As such, it should be understandable by a broad range of stakeholders, including managers, engineers, policymakers, regulators, academics and a lay audience. To achieve these objectives, the framework is simple, but further detail is included where it is required for the overall objectives of this study.

Ultimately the framework should provide a structure that facilitates:

- analysis of interactions between the layers to better understand the consequences of governance and regulation on management procedures, through technical and task-level issues, to risks associated with the system
- evaluation of existing approaches for safety analysis and managing risks at each layer
- development of maturity models to better understand, for example, regional and sector-specific differences in managing the risk of complex systems
- identification of potential approaches to taking a holistic view of risk management of complex systems across the layers that avoids 'not seeing the woods for the trees' [14] from any particular perspective.

At this stage, the framework provides a structure that should give a basis for meeting these requirements, but it is not yet sufficiently mature to meet all of them. It should also be seen as the first iteration of the framework that we would expect to be refined in the remainder of the **Safer Complex Systems** programme.

The framework provides conceptual clarity around the factors that lead to systemic failures that have a safety impact in complex systems. It is the first iteration of the framework that will be refined in the remainder of the Safer Complex Systems programme.

# 3.2 The framework

The framework has six major elements, as identified in Figure 2. The individual elements are described below, supported by more detailed definitions of their constituent factors and references to source material in Appendix A.3. A wide range of examples that motivated the development of the framework and that show its use to describe systems and situations are given in Appendix C, with one illustrative example given in Section 3.3.

The framework, at this stage in its development, seeks to provide an accessible overview of the elements and factors that influence the safety of complex systems. As presented, it indicates only the highest-level dependencies between the elements. This is adequate for descriptive purposes but, to be used analytically, the framework would need to be expanded to include more explicit links between the elements and factors (see Section 3.5). In contrast, showing the interdependencies explicitly would run counter to making the framework accessible and using it for descriptive purposes.

As visualised in Figure 2, the central axis of the framework shows a flow from causes of system complexity via their consequences to systemic failure. This is similar to the progression from faults due to errors to failures underlying traditional functional safety engineering. However, it should be noted that systemic failures arise out of complexity, not from the faults in system elements, and that the interdependencies between elements and factors are more subtle than a simple cause-effect relationship.

The emergence of systemic failures can be tempered by actions at design-time and during operations. These risk management controls should reduce the likelihood that systemic failures arise. However, the framework also recognises exacerbating factors that can make systemic failure more likely, which are comparable to common cause failures in traditional safety engineering.

Each of the six top-level framework elements are first defined more formally, then broken down into their constituent *factors*, reflecting the information gained from stakeholder consultation, from the literature, and from analysing the case studies including accidents and incidents (see Appendix C).

The framework has six major elements and shows a flow from causes of system complexity via their consequences to systemic failure. It seeks to provide an accessible overview of the elements and factors that influence the safety of complex systems.

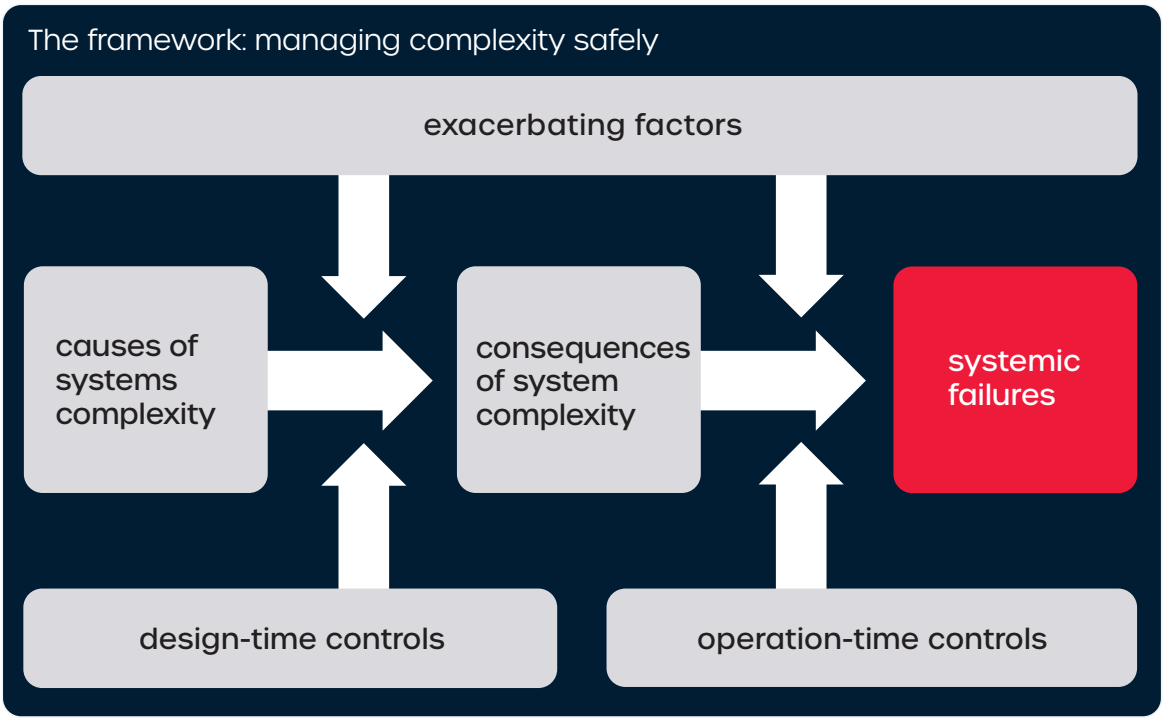



Figure 2: The Framework: Managing complexity safely

*Framework elements*

The framework elements are as follows, starting with the middle causal flow and ending with the exacerbating factors:

- **Causes of systems complexity** – factors at the governance, management and task and technical levels that engender complexity in systems, building on the concepts from complex systems theory, for example rapid technological change.
- **Consequences of systems complexity** – manifestations of complexity at the governance, management and task and technical levels, which can lead to unsafe behaviour if not properly managed, such as unintentional and unrecognised risk transference between stakeholders.
- **Systemic failures** – failures relating to the whole system, rather than a particular part, that impact the safety of some or all of the stakeholders in a system, for example inadequate regulatory control.
- **Design-time controls** – approaches that can be applied at the governance, management and task and technical levels to reduce the causes of complexity and/or to reduce the likelihood that the consequences will occur, such as inclusive design.
- **Operation-time controls** – approaches that can be applied at the governance, management and task and technical levels to reduce the likelihood of the consequences of complexity giving rise to systemic failures or reducing the risk associated with such failures, for example contingency planning.
- **Exacerbating factors** – things that make the management of complexity more difficult perhaps inhibiting both design time and operational management strategies. This might be conflicting legislative requirements on the system as a whole or between system elements.

Each of these six elements of the framework are described below in the order that they are introduced above. The descriptions are intended to illustrate 'fault propagation' between model layers (see also Figure 10 in Section 4). A definition of all the factors in each element is provided in Appendix A.3 but some illustrative examples are provided here.

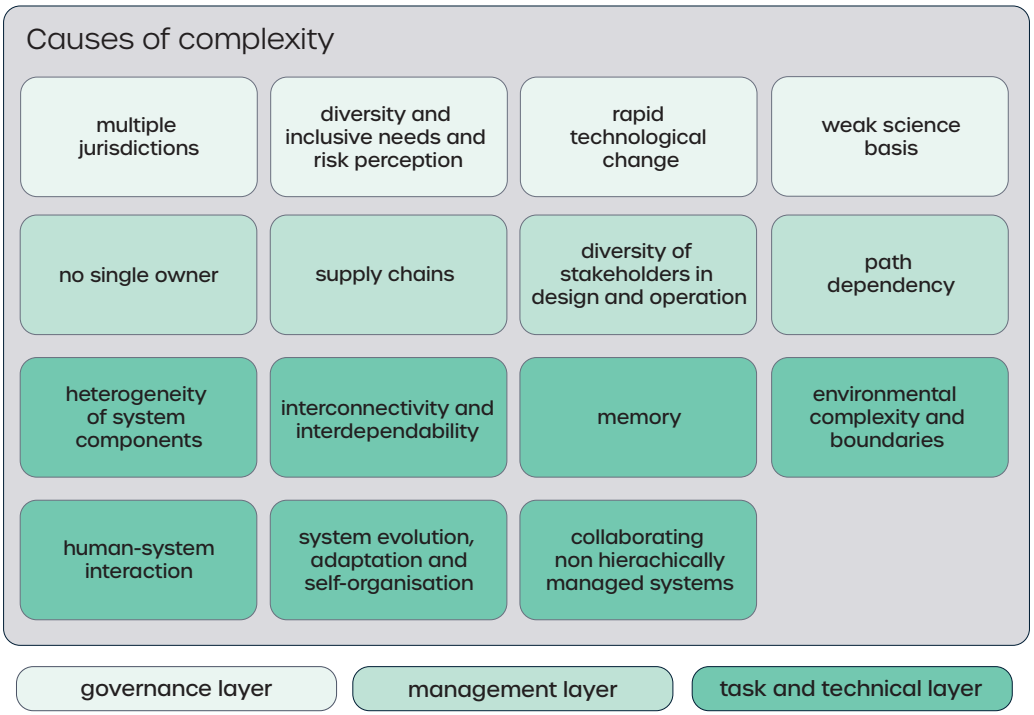



Figure 3: Causes of complexity

### *Causes of systems complexity*

The characteristics of complexity, from the perspective of complex systems theory, are set out in Figure 1 and Appendix A.2. The scope of consideration has been expanded to identify causes of complexity at the governance, management and task and technical layers, re-expressing some of the concepts introduced earlier for ease of understanding. These *factors* are summarised in Figure 3 and each of the layers is briefly characterised and illustrated below.

- **Governance** – factors that mean it is unclear how to judge the safety or acceptability of a system, for example autonomous maritime vessels will be subject to regulations in multiple jurisdictions especially if they can be operated normally (current regulations agreed through the International Maritime Organisation (IMO) apply), fully autonomously (as yet no regulations, or differing regulations for individual local jurisdictions) or remotely operated (where some regulations for land-based facilities will apply).
- **Management** – factors that, in the main, introduce uncertainty into the responsibility and accountability for actions. For example, path dependency is apparent in the NATS failure, see Section 5.1, where legacy software components 'failed' when exposed to a novel system configuration.
- **Task and technical** – factors in the complexity of the system itself, the complexity of the environment, and of human-system interactions. For example, environmental complexity and open system boundaries is apparent in the Uber Tempe fatality, see Section 5.2, where the vehicle failed to recognise a pedestrian pushing a bicycle.

As an illustration of the propagation model set out in Figure 10, the consequences of a weak science base, shown in Figure 3, are competency gaps, standards and regulations lag, see Figure 4. As a concrete example, some industries and professional bodies are now beginning to investigate issues of artificial intelligence (AI) and machine learning (ML) [15].

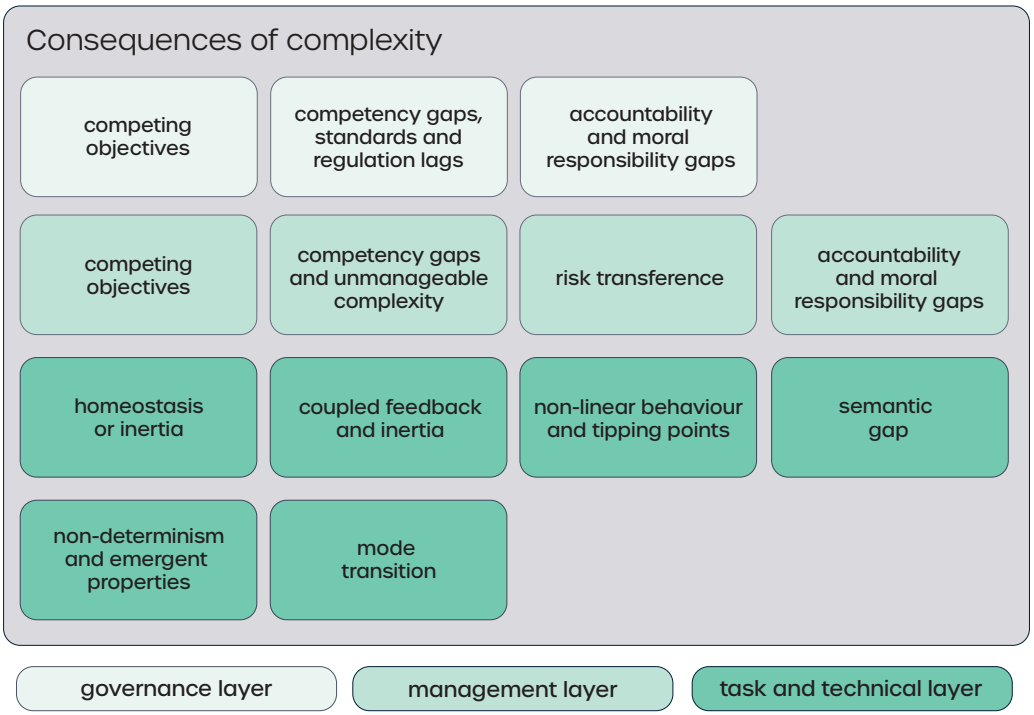



Figure 4: Consequences of complexity

### *Consequences of systems complexity*

The consequences of complexity extracted from the information gathering, stakeholder consultation and case studies are summarised in Figure 4 and each of the layers is briefly characterised and illustrated below.

- **Governance** – uncertainty in apportioning responsibility and accountability for, or determining expectations on, system and organisational behaviour, for example the absence of regulations for safety operating systems such as balloons and aircraft in the stratosphere is an example of competency gaps, standards and regulations lag.
- **Management** – difficulty in ensuring equitable exposure to safety risk and in responsibility and accountability for those risks, for example the risk transference to the pedestrian in the Uber Tempe fatality (see Section 5.2), where Uber were found to have no case to answer by the State of Arizona.
- **Task and technical** – behaviour at variance with expectations of safety in credible but unanticipated operational circumstances, for example the delay in diagnosing or recognising sepsis in the fatality in a Galway hospital, see Appendix C.3.1.

The example of the E. coli outbreak in Germany (see Appendix C.4.1), where competency gaps, standards and regulations lag, see Figure 4, led to the systemic failure of inadequate regulatory control, see Figure 5. This contributed to infected beansprouts contaminating the supply chain, despite them having undergone the required tests, and again illustrates the propagation model; see Figure 10.

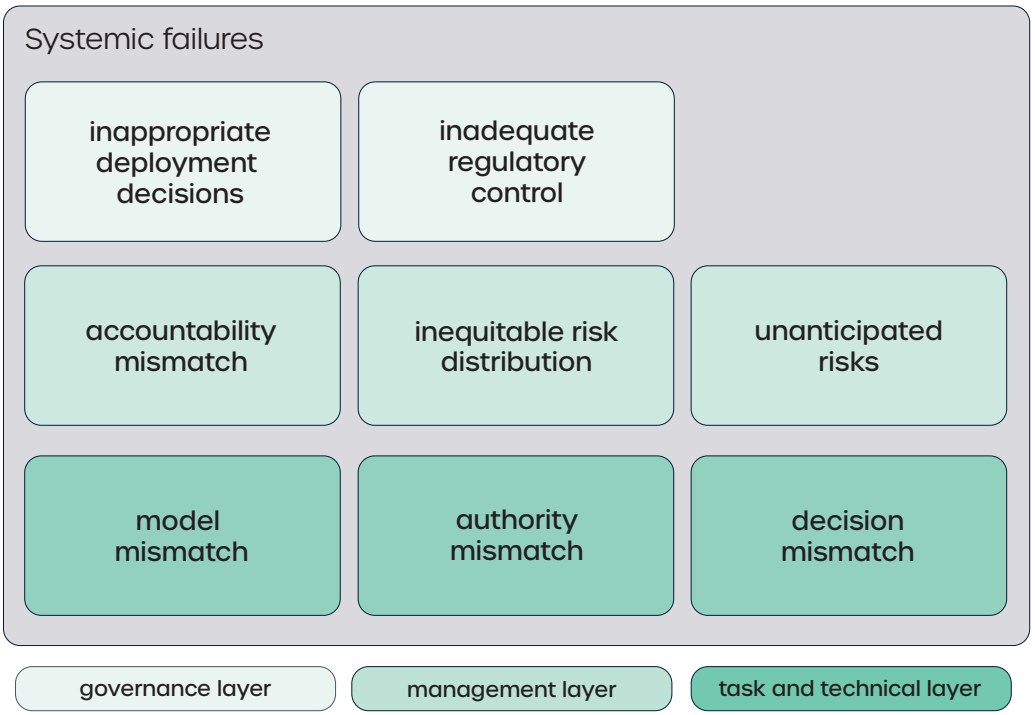



Figure 5: Systemic failures (initial model)

### *Systemic failures*

The categorisation of systemic failures of complexity was motivated in Section 2.2. Here, systemic failures that have been identified in the study to date are summarised in Figure 5 and each of the layers is briefly characterised and illustrated below. This is the most novel area of the framework and we are not aware of any other attempts to characterise systemic failures, so this is likely to evolve in later phases of the **Safer Complex Systems** programme.

- **Governance** – in essence these are situations where systems are deployed that shouldn't be (they are unsafe) or the controls are insufficient to ensure continuing safety. For example, based on the Congressional hearings the Federal Aviation Administration's (FAA) decision to allow the Boeing 737 MAX to continue flying after the first accident could be viewed as an inappropriate regulatory control (see Appendix C.1.2).
- **Management** – inappropriate risk imposition, including lack of accountability for risks. For example, the inequitable risk distribution, putting the mother at risk when the loss of her baby was inevitable in the fatality in a Galway hospital (see Appendix C.3.1).
- **Task and technical** – in general, mismatches between the intended behaviour of the system or humans interacting with the system. An example of this is the inability (extreme difficulty) of the pilots over-riding the Maneuvering Characteristics Augmentation System (MCAS) system on the Boeing 737 MAX could be viewed as an authority mismatch (see Appendix C.1.2).

In terms of the propagation model, inadequate regulatory control can give rise to, among other things, problems in supply chains and cross-domain collaboration, for example, failing to control the provenance of replacement parts when maintaining systems.

It is also important to be aware that systemic failures are not always total, affecting all stakeholders across all levels. The system could fail catastrophically to meet its objective but only for a subset of stakeholders. Or the failure may be felt disproportionately by certain groups, potentially allowing the failure to go unnoticed by those responsible for the system. This form of discrimination raises ethical concerns that organisations need to be aware of as a potential unintended consequence of their operations. It is also a wider issue for society as a source of discrimination against minority or otherwise disadvantaged groups.

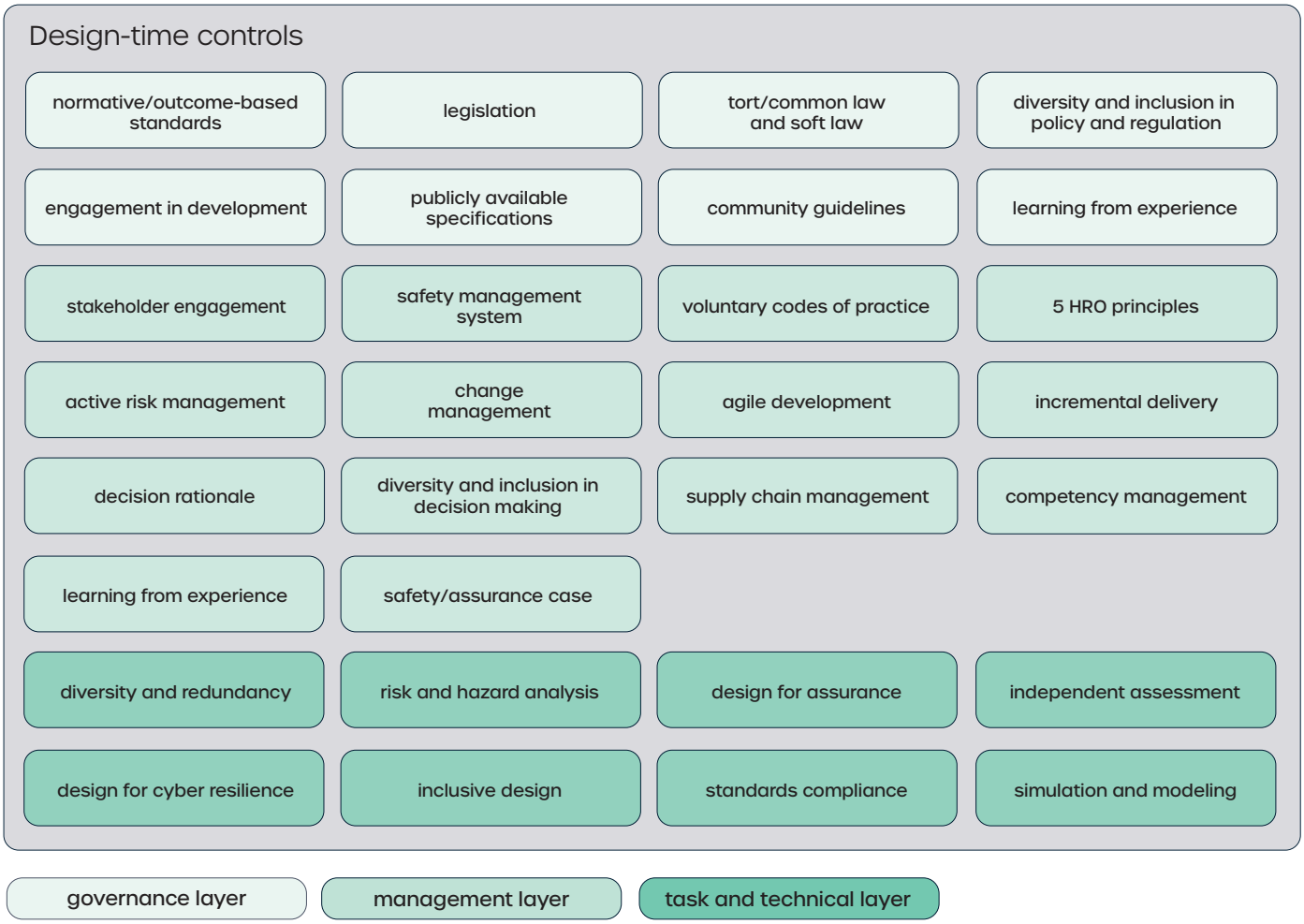



Figure 6: Design-time controls

### *Design-time controls for managing complexity safely*

Traditional safety engineering emphasises design-time controls for 'engineering out' or reducing safety risk. For complex systems this remains important, although there is a need for greater emphasis on operational controls. Further, established approaches from safety engineering are not sufficient for dealing with complex systems, see Section 4, and work will be needed to enhance traditional methods and to develop new ones, especially for agile analysis of ad hoc systems. The design time controls identified in the study are summarised in Figure 6, and then each of the layers is briefly characterised and illustrated.

- **Governance** – industry-wide regulatory mechanisms for controlling safety risk, recognising the rate at which technology and systems are evolving. For example, community guidelines and 'soft law' are more agile than the normal regulatory and standardisation processes and therefore more likely to keep pace with the evolution of complex systems.
- **Management** – techniques for risk management and control that can be applied at the level of an organisation (and supply networks) again recognising the need for agility. Agile development practices are used successfully in some domains, for example [16].
- **Task and technical** – these are classical safety engineering and functional safety activities, grouped together into major controls, for example independent assurance assists with a risk reduction measure known as fault removal (see Section 4). However, we see the need for a greater focus on measures to increase the robustness and resilience of systems to cope with both predictable and unknown anomalous emergent behaviour, including that caused by interactions with operators and users of the systems.

At the governance level, standards organisations recognise the need for greater agility and are promoting the development of Publicly Available Specifications (PAS) for autonomous vehicles [17]. The Global Mining Guidelines Group are producing pan-industry guidelines on safety of modern mining and quarrying systems [18].

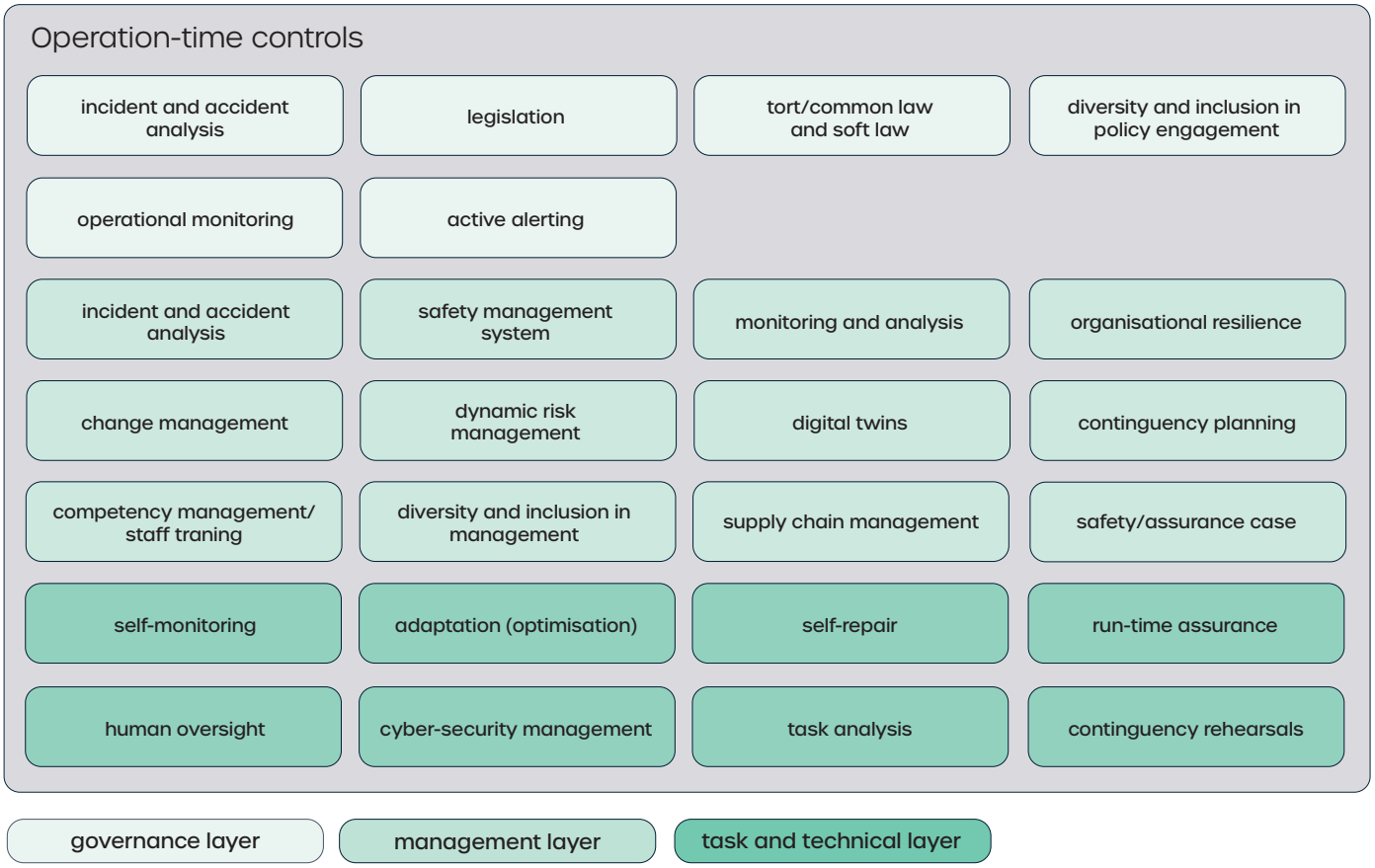



Figure 7: Operation-time controls

### *Operation-time controls for managing complexity safely*

Operation-time controls for the safety of complex systems have a higher weight than for more traditional systems, as emergent behaviour needs to be managed operationally. Furthermore, for systems that were not designed as a whole – often characterised as ad hoc or 'accidental' systems – the only option is operational controls. Thus, there is a need at the management and governance layers to recognise when a new system emerges that needs to be managed; such 'triggers' are outside the scope of the discussion here but will be important in an overall process for managing safety of complex systems (see the discussion in Section 6.5). The controls identified in the study are summarised in Figure 7, and then each of the layers is briefly characterised and illustrated. As with the design-time controls, work will be needed to enhance existing practices and to develop new ones to deal with emerging complexities – advances will be needed across all three layers, perhaps most at the management and governance layers, as more systems have global reach or impact. As noted earlier, where there is a design phase, the design and operational activities will overlap in time and the operational controls may need to be enabled by the design.

- **Governance** – mechanisms for understanding, communicating and responding to risks for the whole domain, for example using information from one system to prompt investigation and/ or remedial actions for other systems.
- **Management** – organisation-level mechanisms for understanding, communicating and responding to system risk, for example analysing operational data through digital twins to identify leading indicators of risks to prompt action to avoid accidents.
- **Task and technical** – mechanisms that implement, or support, active risk management as the system operates, for example rehearsing for contingencies to improve the ability to deal with emergencies should they arise and increasing the transparency of the system behaviour to its users.

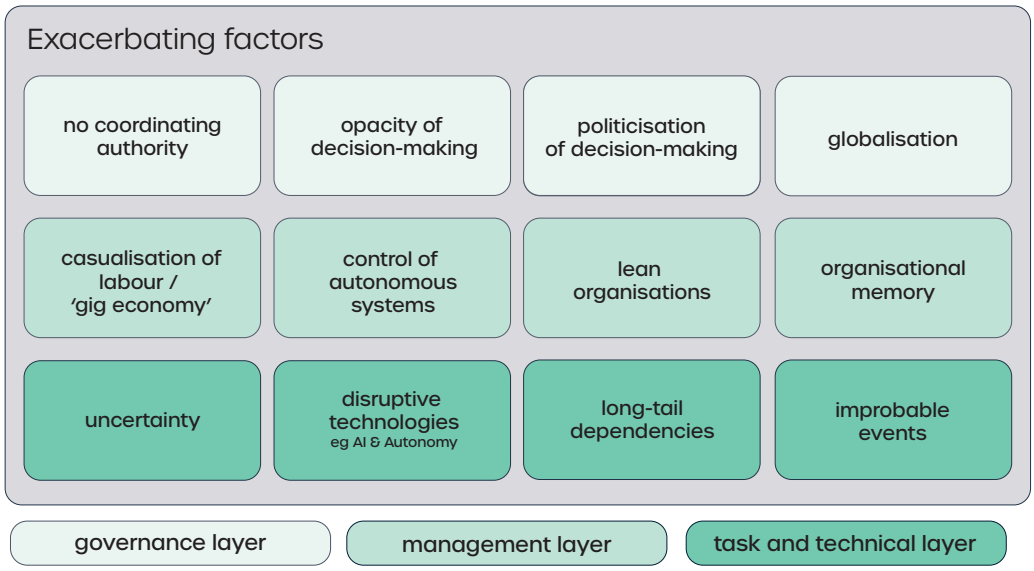



Figure 8: Exacerbating factors

Many of the controls, especially at the governance and managerial levels reflect good practice in dealing with today's complex systems, which have proven successful in some domains (see Section 6.3). Much of this work is brought together under the heading of 'resilience'. Good practice and guidance on resilience for cities and local authorities in the UK can be found at [19] in response to the Civil Contingencies Act (CCA) and a very broad-based set of academic resources can be found at [20]. However, it is anticipated that additional operation-time controls will be needed for complex systems and these are discussed in Sections 6 and 7. As mentioned above, there are links between design-time and operational controls – for example, human oversight can be enabled by inclusive design.

### *Exacerbating factors*

The exacerbating factors make management of safety in complex systems worse, either by 'amplifying' the causes of complexity or by limiting or impeding the control strategies. These are similar to, but broader than, the concept of escalating factors used in Bow-Tie Diagrams (BTDs) [21].

- **Governance** – these are mainly factors that inhibit international collaboration and coordination in situations where consistent global responses are needed. For example, the absence of a global coordinating body (or having such a body that is ineffective) can lead to different regulations in different jurisdictions, thus different approaches to manage systems, and the need for different technical solutions for systems to work in multiple countries.
- **Management** – these are mainly 'people-based' factors that can inhibit achieving effective safety management, or a good safety culture. The 'gig economy' in the UK means that it is quite common for many of the workers on a building site to be self-employed, which makes it harder for the main contractor to instill a coherent approach to safety and a positive safety culture.
- **Task and technical** – these *factors* are mainly concerned with uncertainty and low-probability, high impact events that are hard to address at design time and which are also hard to assure and manage. For example, events that were thought to be independent prove to be correlated making their management more challenging; again human factors are important here as humans are often able to adapt to unforeseen circumstances and thus are instrumental in achieving resilience.

As an example of the impact of lack of international coordination, Eurostar trains carry multiple signalling systems (Belgian, British and French) to enable them to operate in all three countries.

# 3.3 Illustration of the framework

To make the role of the framework more explicit, it is illustrated by considering how to describe the complex interactions in the supply network of personal protective equipment (PPE) used in healthcare, specifically during the COVID-19 pandemic.

It is necessarily limited in scope and focuses on PPE supply in the first three months of the pandemic, in the UK and the USA, and does not purport to be a complete analysis of the management of COVID-19.

The illustration in the box opposite shows links across the layers and between domains.

## Illustration of the Framework – Supply of PPE during the COVID-19 pandemic

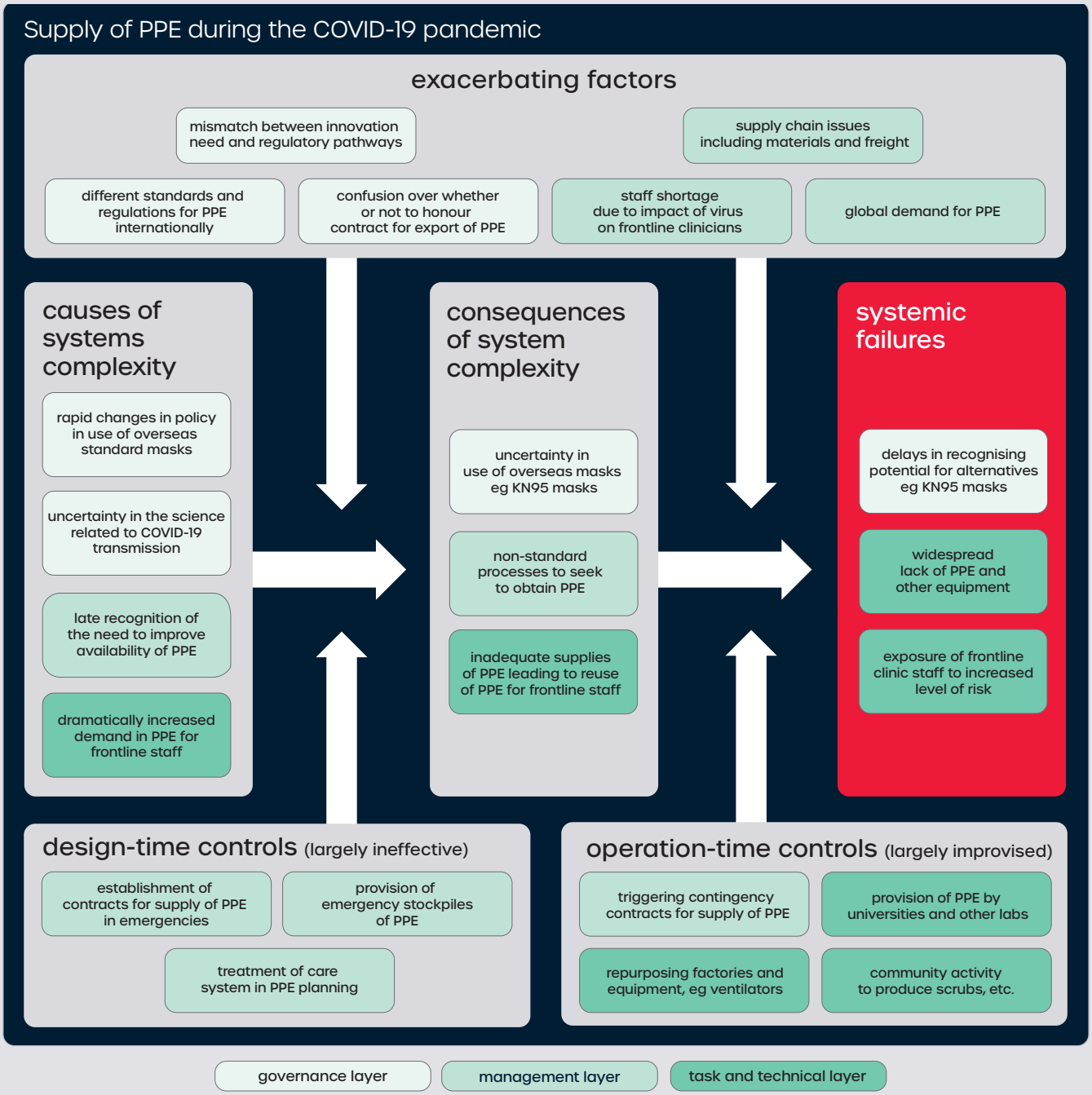
*The views and opinions expressed in this case study are those of the University of York research team and do not necessarily reflect the views of Engineering X*

**What happened?**
The COVID-19 pandemic started (was first recognised) in Wuhan, China at the end of 2019 and has become a pandemic affecting most countries in the world. It is still ongoing at the time of writing. The global healthcare enterprise and its response to COVID-19 can be viewed as a system. However, it is too complex to cover here and instead this example focuses on the supply of PPE, gowns (known as 'scrubs') and other specialist equipment such as ventilators – a supply network system. In the UK and the US there have been significant shortages of such equipment and there have been a range of formal attempts by hospital management, responses by businesses and 'community' activities to respond to the issues.

**Why did it happen?**
The figure below shows how the framework can be used to illustrate PPE-related systemic failures in the supply network. There is a large, fast-changing set of (media) reports on the issues, so it was only possible to include a limited number of references. The most commonly used standard for respiratory PPE (masks) is N95. However, there is a Chinese equivalent known as KN95 and there were (rapid) changes in US government policy regarding acceptability of KN95 masks, from rejection [22] to acceptance [23]. There was also confusion about whether or not contracts to export masks should be honoured [24]. Although it is common practice to have significant stocks of PPE, in some cases these stocks had been allowed to run down [25]. In the UK there were also failures to include the core system in PPE planning.

Once it was recognised that there was a growing need for PPE, contingency contracts were invoked. But in some cases these involved overseas suppliers who were more concerned about national needs (or governments imposed controls) so the demand was not easily met. This led to some unusual tactics to obtain masks including a senior clinician personally inspecting PPE and ensuring that it wasn't appropriated by Federal officials [26]. There was also a widespread move to support the PPE requirements, with universities and other laboratories donating equipment [27], companies re-purposing factories and hospitals adapting to the situation, for example by reusing PPE. There are also more social-level activities, such as individuals who were unable to undertake their normal work making good use of their time by producing scrubs for hospitals [28]. This example shows both unexpected interdependencies between systems, and also the role of creativity and flexibility in resilience.

The problems were exacerbated by variations in standards, the mismatch in pace between the need for innovation and regulation (although some rapid approvals were given) and supply chain issues including availability of materials and problems of transport (freight). Uncertainty in the science relating to the transmission of the virus is shown as a causal factor, although it could also be viewed as an exacerbating factor.


Supply of PPE during the COVID-19 pandemic
exacerbating factors
mismatch between innovation need and regulatory pathways
supply chain issues including materials and freight
different standards and regulations for PPE internationally
confusion over whether or not to honour contract for export of PPE
staff shortage due to impact of virus on frontline clinicians
global demand for PPE
causes of systems complexity
rapid changes in policy in use of overseas standard masks
uncertainty in the science related to COVID-19 transmission
late recognition of the need to improve availability of PPE
dramatically increased demand in PPE for frontline staff
consequences of system complexity
uncertainty in use of overseas masks eg KN95 masks
non-standard processes to seek to obtain PPE
inadequate supplies of PPE leading to reuse of PPE for frontline staff
systemic failures
delays in recognising potential for alternatives eg KN95 masks
widespread lack of PPE and other equipment
exposure of frontline clinic staff to increased level of risk
design-time controls (largely ineffective)
establishment of contracts for supply of PPE in emergencies
provision of emergency stockpiles of PPE
treatment of care system in PPE planning
operation-time controls (largely improvised)
triggering contingency contracts for supply of PPE
provision of PPE by universities and other labs
repurposing factories and equipment, eg ventilators
community activity to produce scrubs, etc.
governance layer
management layer
task and technical layer

# 3.4 Cross-cutting topics

There are a number of cross-cutting issues that affect the **Safer Complex Systems** programme as a whole and that need to be reflected in the use and evolution of the framework. They are briefly outlined here, but all require further study.

### 3.4.1 Internationalisation

Many systems, such as civil air transport, have an international reach, and in other cases (essentially) the same systems are deployed worldwide. In some industries, like aerospace and maritime, there are global organisations that coordinate standards and approaches to regulation – the International Civil Aviation Organisation (ICAO) and International Maritime Organisation (IMO), respectively. While the details are complex, these international bodies do ensure a level of consistency across nations, but they tend to move slowly – with standards and regulations lag as a consequence. However, not all industries have such bodies and, even where they do exist, globalisation is not without its challenges. Some key issues are [29]:

- **Risk acceptance** – different risk acceptance regimes are used in different countries (see also Section 4.2), so product designs may need to be modified for different geographies, or made more complex to be truly global products.
- **Value for the prevention of a fatality (VPF)** – many risk acceptance regimes include cost considerations, implicitly or explicitly basing decisions on VPF (the investment that should be made to save a single statistical life – not what a person is 'worth') and this value can vary substantially with the wealth of nations.
- **Standards** – in many jurisdictions, international standards take precedence over national ones, but in many cases there are still 'local' standards to adhere to – this is very evident, for example, in electrical plugs and sockets, which do reflect different attitudes to electrical safety.
- **Professional competency** – educational norms and professional standards vary from country to country, and there are few (internationally) recognised qualifications in safety; further, some of those that are recognised, such as TUV Rheinland [30], tend to be based around standards compliance and not the broader issues identified in this study.

There are good examples of community guidelines in some sectors where there are no international bodies. For example, the Global Mining Guidelines Group, a collaboration between equipment manufacturers, site operators and regional authorities, has produced guidelines on autonomous systems in mining and quarrying [18]. Nonetheless, globalisation remains an issue that will require attention in the remainder of the **Safer Complex Systems** programme.

### 3.4.2 Equality, diversity and inclusion

Diversity provides mechanisms for achieving improved results in safety performance but it also presents challenges for the operation of complex systems. In some situations the term heterogeneity is used or preferred to diversity as it is less value laden.

*Challenges*

One of the greatest challenges for achieving and assuring safety of complex systems is heterogeneity in risk perception. This can exist within or between societies and is particularly evident at the governance and management layers. The difficulty in agreeing acceptable levels of safety makes implementing risk controls more challenging. This can be particularly true for systems and services used across international boundaries and social-economic groups. Some societal groups will be impacted more than others by the negative effects of unsafe complex systems, but they or others may also be (disproportionately) impacted by the cost of effective control implementation.

Mechanisms are needed to explicitly consider, assess and address the impact of heterogeneity in risk perception and management, both at the governance and management layers.

*Benefits of diversity and inclusion*

Inclusion engages individuals and values everyone as being essential to the success of an organisation. Such a cultural approach is linked to higher-performing organisations [31].

Implementation of appropriate equality, diversity and inclusion practices is critical to achieving an appropriate safety culture within or across organisations. In particular, the following inclusion-based behaviours play a significant role in safety culture:

- **Open and trusting environment** – an open and trusting environment with an absence of prejudice and discrimination and where everyone is respected and valued.
- **Devolved decision-making** – decision-making processes that are devolved to the lowest point possible.
- **Listening, encouragement, participation, honest engagement –** the encouragement of consultation and participation, with management actively listening to and acting upon what employees are saying.
- **Understanding of core values** – an understanding of core values by all stakeholders.
- **Open flow of information** –

an open flow of information throughout the whole organisation between all levels, so that business goals are communicated to everyone, and an attitude of 'us and them' (employees and management) is discouraged.

- **Innovation/creativity** – the encouragement of innovation and creativity.

Cognitive diversity is a critical component of controls to mitigate the impacts of complexity at both the governance and management layers, and at design-time and operation-time. The impact of emergent behaviour of complex systems and the exacerbating factors identified in this report highlights the importance of leveraging heterogeneous perspectives for tackling these challenges.

Some of the case studies show where it may have been possible to achieve safer outcomes if the principles of equality, diversity and inclusion had been considered. For example, in the Boeing 737 MAX example, a culture of concealment, as identified by the House Committee, directly conflicts with an open and trusting environment (see Appendix C.1.2). Further, there may be benefits in accident investigation embracing equality, diversity and inclusion.

A lack of diversity across all layers also increases the possibility that important aspects will simply be missed due to unconscious bias or the required perspectives not being present in the analysis, which could end up being a cause of systemic failure as the developed solution would be found to be unsuitable when applied in a particular domain. Systems thinking and complexity thinking can help mitigate this as they are inherently interdisciplinary approaches to solving problems and seek to draw input from diverse domain experts and other stakeholders to see the connections and inter-relationships present in the system being studied.

### 2.3.3 AI and autonomy

The increasing use of artificial intelligence (AI) techniques, in particular machine learning (ML) approaches, such as deep neural networks [32], to automate key tasks of safety-critical systems raises extremely challenging questions when reasoning about the overall safety of the system. Use cases already under development include detection of obstacles in the path of autonomous vehicles [33] and decision-making support in treating sepsis patients [34].

Increased automation within a complex system, regardless of its technical implementation, can be reflected in many facets of the framework ranging from the allocation of responsibility in case of failure (who is to blame: the manufacturer, operator or user?) through to the selection of appropriate design and operation time controls. AI, even when used for very restricted functionality within the overall system, also exhibits many of the characteristics of complex systems. Complexity and uncertainty in this context can be understood in terms of the unpredictability of the system's operational domain; the complexity and unpredictability of the system itself; and the increasing transfer of decision-making function from human actors to the system. The framework outlined in this report is intended to provide a means by which the challenges associated with AI and autonomy in safety-critical systems can be systematically analysed from a multi-disciplinary perspective. Some of the most salient issues can be summarised as follows:

- **Governance layer:** The use of AI in autonomy leads to a number of ethical questions [35] related to the transfer of decision authority to systems whose actions cannot be easily explained, leading to accountability and moral responsibility gaps [1]. These include issues related to bias and discrimination as well as expected behaviour in inherently hazardous and ambiguous situations. Appropriate forms of regulation, based on a thorough understanding of the potential and limits of the technology, need to be developed alongside broad public consultation and appropriate legal frameworks.
- **Management layer:** Safety engineering methods and safety management systems must be adapted in order to incorporate new classes of risks introduced by AI and ML. Initial work on developing assurance arguments for such systems has begun [36], but there is no clear consensus in the form of standards or regulations. ML techniques tend to be highly sensitive to the context in which they were trained and the training data selected from within this context. Manufacturers of such systems must therefore provide clear guidance on the limits to the operational domains for which the systems can be considered safe and also continuously monitor the use of the systems to discover whether or not they respect these operational domains.
- **Task and technical layer:** AI algorithms and ML in particular deliver inherently uncertain results. For a given video frame they might classify the probability of a pedestrian inhabiting a certain portion of the picture as 83%. But in the very next frame, imperceptibly different to the last, they may misclassify the same object as only 26% pedestrian and 67% road sign. Furthermore, the individual decisions made by the algorithms, based on the millions of different weights in the neural network are difficult to explain. This leads to an unclear understanding of their performance. Design-time and

operation time controls are therefore required to reduce and quantify this level of uncertainty as far as possible, as well as to select appropriate measures to ensure robustness at the system-level. The ability to continuously improve the functionality based on updated training data from real-world operations can also be used to increase system resilience, however care must be taken to ensure 'monotonic' safety improvements that guarantee that previously qualified behaviour remains safe despite changes in the underlying model.

The interest in AI for safety-critical systems is not just based on academic curiosity. Evidence suggests that AI systems can often perform better than humans in some situations, thus leading to ultimately safer systems (for example reduced number of traffic accidents, better informed medical decision making). However, there is still much uncertainty surrounding the arguments that such systems are acceptably safe and do not contribute to systemic failure. Further work should apply the framework to analysing in detail the use of AI within safety-critical systems to develop holistic approaches to ensuring safety across the governance, management and task and technical layers, despite the inherent challenges in creating convincing assurance arguments for such systems.

The increasing use of AI in safety-critical systems raises extremely challenging questions about the system's overall safety due to the openness of the system's operational domain; the complexity and unpredictability of the system itself; and the increasing transfer of decision-making from human actors to the system.

# 3.5 Maturity and evolution of the framework

The framework presented here is an initial version, which we expect to be refined and validated during the remainder of the **Safer Complex Systems** programme. Such a framework can be viewed as evolving through three broad phases:

- **Descriptive** – can be used to describe systems and situations in a clear way, providing a basis for a learning from experience and improving designs, such as systematic classification of the causes of an accident.
- **Analytical** – can be used to assess properties of the system or situation prior to operation, for example to predict operational risk or risk distributions, and to assist in operational monitoring. This will involve integrating layer and domain specific analysis methods and models such as SEIPS for the healthcare domain [37]. These models and analytical tools should help in choosing design time and operational controls, including informing trade-offs. They should inform a more holistic view of the system taking into account aspects of complexity and resulting potential systemic failures.
- **Generative** – can be used to produce parts of the system or, more likely in this case, producing associated information, for example automatically producing fault trees or (parts of) a system safety case.

At present the framework is descriptive, as illustrated by the case studies in Appendix C. The framework can usefully be expanded to include other factors and safety management controls to make it a richer descriptive tool. This can be investigated in the second phase of the **Safer Complex Systems** programme, for example by testing it against additional case studies (see Section 7.3).

To be used analytically, the framework needs to be enhanced and underpinned by models. These enhancements need to be investigated in later phases of the programme, but some key steps can be identified here.

First, the controls can be correlated with the causes and consequences of complexity – identifying where they are appropriate and effective. This would help to establish criteria for selecting design-time and operation-time safety management controls appropriate for the system and situation. The discussions in Section 4.5 give an illustration of what needs to be done in the area of safety and risk analysis methods, but doing this more broadly is likely to require an extensive research activity. Also, producing the correlation of factors and controls should identify those factors without known controls, which therefore require investigation that might stimulate a wider research activity.

Second, the controls can be stratified, identifying the most basic controls (those to introduce first), controls that can build on the basic stratum, and so on – potentially producing a maturity model for the controls as has been proposed for quantitative risk analysis [38]. This clearly would need to build on the work in correlating complexity factors and controls. There would be benefit in focusing on this in the next phase of the study as it will help to make the framework more readily applicable in a range of situations.

Third, the framework would benefit from mathematical methods or data analysis tools that enable actual and proposed systems or situations to be analysed. Some of the long-established work on theory of complex systems is relevant here; see also the discussion of supply networks in Appendix C.4. As noted above, many modern complex systems are data rich and there is an opportunity to carry out analysis and build digital twins for safe exploration of design alternatives; these will become a key part of an analytical framework (see Section 7.1.4).

Perhaps the most important area is to investigate and include more systemic failure classes. Many methods in systems safety engineering and functional safety use classifications of failure modes to guide analysis, for example the omission, commission or incorrect provision of a function, see Section 2.2. Enriching the framework so that it could be used in support of analysis at the causal level (cause– consequence–systemic failure) would include identifying appropriate classification of the system failure modes. For example, for model mismatch these might include:

- **Framing** – information about the real-world needed by the system to make safe decisions is not included in the system's model.
- **Timing** – the system model lags the real-world data sufficiently that decision-making is unsafe, for example, a decision that was safe based on the model is not safe at the time the decision was made and implemented.
- **Externally inconsistent** – the model is (significantly) at variance with models used by other systems in a system of systems, or by operators, leading to dangerously inconsistent decision-making (for example see the Watchkeeper example in Appendix C.1.7).

These should be seen as illustrative examples, and a proper study is required, but such classification could form a key part of a HAZOP-like hazard identification method. Such ideas could initially be addressed in the second phase of the study, but they are likely to require a more extensive programme of research.

Developing the framework to be

generative is a much longer-term activity. There are generative safety analysis tools, for example, some model-based safety analysis (MBSA) tools can generate fault trees from system models enhanced with information about fault propagation [39, 40]. There has been work on MBSA for some years, and there are industrial applications of the technology, such as in the aerospace and automotive domains. More recently, there has been work on generative safety and assurance cases, partially deriving the safety case from failure models, for example BTDs, using safety argument patterns [41]. Such work is likely to be necessary to fully support dynamic risk management as one of the operation-time safety management strategies, but cannot be addressed fully until progress has been made at the analytical level, so this too is a longer-term research issue.

Considerable work has been done in systems engineering to produce modelling notations and supporting tool-sets, such as SysML [42] and SPES [43]. These notations typically support multiple views, for example physical and functional decompositions, and often are quite wide-spectrum covering aspects of the environment and stakeholders, although they typically don't support a safety viewpoint. Some of the work on MBSA has also built on widely used models such as Systems Modelling Language (SysML), for example to generate fault trees for aero engine control system models [44]. The use of such notations is likely to be beneficial in making the framework both analytical and generative, and they should be studied, at least from this point of view, in later phases of the **Safer Complex Systems** programme.

The framework is an initial version, which will evolve so that it can in time be used to analyse systems and help to generate associated information, for a system safety case for example.

# 4

# Safety analysis and management techniques

The aim of this section is to consider how we define and measure risk and identify safety analysis and management methodologies and tools suitable for the complex and interdependent systems in our society today. While the report does not attempt to produce a definitive list of the most appropriate methods, it does identify the need for an approach for selecting appropriate combinations of methods for application at and across the governance, management and task and technical layers.

# 4.1 Introduction

The underlying objective of the **Safer Complex Systems** programme is to identify measures for managing risk and increasing safety by improving the design, management and governance of complex systems. A key component of such a framework must therefore be the ability to evaluate the safety risk associated with the system as well as the impact of any proposed measures for reducing risk. Much work has already been done on developing various risk and safety analysis techniques for engineered systems, some examples of which will be described below. However, this study focused specifically on how aspects of complexity influence safety and how such risks can be managed.

Risk can be defined as the extent of a system's inability to meet its (safety) objectives and is often expressed in terms of the product of the probability of a system failure and the severity of its consequences. In the context of this study, these consequences are interpreted to be failures of the system to fulfill its objectives (see Section 2.2). Such failures may include single catastrophic events such as loss of an aircraft, but also an erosion of safety properties over time, such as omitting to protect the public and the environment from harmful side-effects of widely used chemicals. Laprie et al [45] defined a model of how faults in individual system components cause an erroneous system state (error) that may subsequently lead to users experiencing a failure of the system's service. Furthermore, for systems consisting of many independent interacting components or sub-systems, a failure at a particular level within a system hierarchy or causal chain could manifest itself as a fault at the next higher level or dependent system (see Figure 9). Causal approaches to modelling fault propagation form the basis of many widely used safety analysis measures. These measures can be classified as follows [45]:

- **Fault prevention** – ability to prevent the occurrence or introduction of faults.
- **Fault tolerance** – ability to avoid service failures in the presence of faults.
- **Fault removal** – ability to reduce the number and severity of faults.
- **Fault forecasting** – ability to estimate the present number, the future incidence, and the likely consequences of faults.

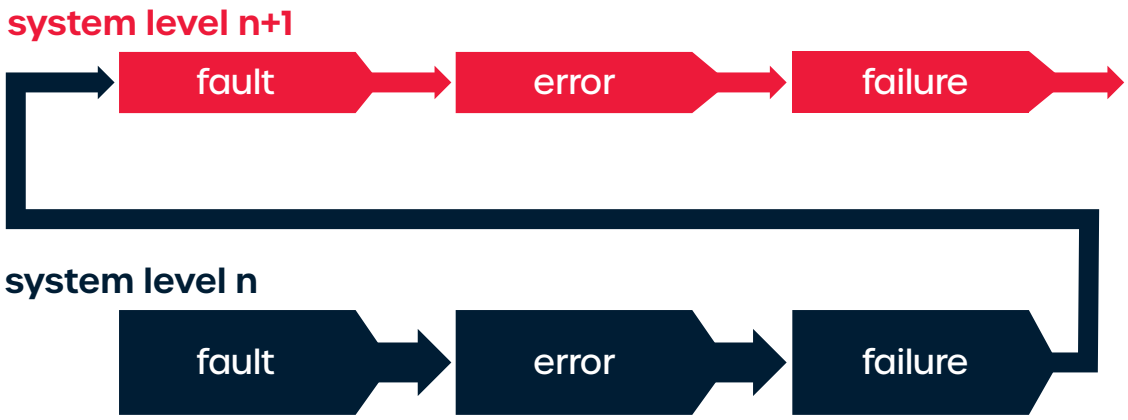



Figure 9: Fault propagation

## 4.2 How safe is safe enough? Measuring risk and setting targets

The starting point of any safety argument is some definition of the safety claim that is being made. In other words, how safe do we (need to) argue the system to be? In terms of safety analysis and management techniques, this involves setting some targets to compare to the postulated or achieved level of risk imposed by the system.

Risk can be assessed based on quantitative measures such as fatal accident rates, probability of individual component faults and a system's impact on overall mortality rates. In 2016, the German Ministry of Transport and Digital Infrastructure commissioned a report [46] into ethical considerations of automated driving. One of the recommendations of the report was that it must be shown that the automated driving systems perform, on average, better than a human driver in terms of avoiding or mitigating hazardous situations. A related approach is the French *"Globalement au moins aussi bon"* (globally at least as good), or GAMAB. Here, the principle is that any new system must be at least as good as any previous system it replaces. The difficulties of such quantitative approaches to defining acceptable levels of risk include the choice of representative target values to make a fair comparison as well as the collection of statistical evidence to back up any claim that the achieved level of risk is sufficiently low. For example, which standard of human driver should be used as a comparison: an expert driver operating in perfect circumstances or the average member of the public with all the usual distractions and human shortcomings? Also, how can a statistically convincing argument be made before release of the system, when this would require driving billions of miles [47], making the system non-viable economically or potentially delaying any safety benefits associated with the introduction of the technology?

The principle of *As low as reasonably practicable* (ALARP), or variants thereof, are often used in the regulation of safety-critical systems. The ALARP approach to risk assessment involves demonstrating that the cost involved in further reducing the risk would be disproportionate to the benefit gained, and this often includes use of a VPF figure for benefit (see Section 3.4). However, these judgements are typically not made solely on the basis of quantitative assessments but also on an understanding of good engineering practice and existing standards. For example, if it could be argued that applying such standards and practices could result in significantly better performance than an average human driver, then a direct comparison to current accident statistics may not be sufficient. In other words, regardless of the statistical probability of a hazardous event happening, we have a duty to minimise risks that could have been avoided by applying rigorous design and operation measures. This apparently open-ended duty is often resolved by appeal to formal standards. For example, several industry safety standards, such as those derived from IEC 61508 [48] for electrical/electronic/programmable electronic system components, apply the principle of safety integrity levels to define a set of measures appropriate to a risk level derived for the system, based on some method of systematic risk analysis. Depending on the safety integrity level, different sets of design, verification and validation measures are suggested.

Due to the inherent uncertainties in defining risk criteria, especially for complex systems, as well as sector-specific conventions and the widely varying differences in risk perception by the different stakeholders, it is usual to apply a range of criteria for defining an acceptable level of risk for a system. Ultimately, risk will always be subjectively perceived through a set of culturally-specific filters. These could include regional as well as domain specific variations that also evolve over time [49]. In order to cater for these inevitable differences in risk perception, a diverse set of both quantitative and qualitative criteria should be selected in close cooperation with system stakeholders, taking into account the many dimensions of diversity. The role of the human participants in complex sociotechnical systems can have a huge impact on the overall safety of the system. A common understanding of risks and inevitable trade-offs is therefore a critical component of any safety culture across all three layers of governance, management and task and technical.

# 4.3 Safety analysis and management methodologies

This section provides some representative examples of safety analysis and management techniques to investigate the extent to which they can be applied to the specific issues related to complex systems. The examples may motivate a set of recommendations of how safety analysis and management can be integrated into the **Safer Complex Systems** framework proposed in this report (see Section 4.4).

- **Deductive and inductive safety analyses:** This set of analysis techniques, most commonly applied at the technical layers of a system, are based on the causal relationships exemplified by Laprie's fault – error – failure model. Fault Tree Analysis (FTA) [50] takes a deductive, top-down approach to identifying combinations of events that could lead to a system failure. Failure Modes Effect Analysis (FMEA) [51] takes an inductive bottom up approach to analysing the impact of individual component faults. Hazard Analysis and Operability (HAZOP) [52] and related approaches focus on the deviations of a system from its design intention, making use of guidewords to distinguish between various modes of failure and to support completeness of analysis. Bow-Tie Diagrams (BTDs) [21] can be seen as a combination of deductive approaches to determining causes of failures and inductive approaches to identifying their consequences. BTDs introduce controls to mitigate both the causes and consequences of failures; this philosophy is reflected in the framework described in this study, although the focus is on mitigating causes and consequences of complexity that might lead to unsafe systemic failures.
- **Cognitive systems engineering:** Rasmussen [14] introduced a risk management framework based on the analysis of variations in behaviour (rather than fault models) across six levels of the sociotechnical systems: government, regulators/associations, company, management, staff, and work. This approach inspired the layered framework applied by this study. The layer of work has been extended with technical considerations of complex systems and provided a causal structure for analysing how factors contributing to system complexity can lead to systemic failures.
- **System-theoretic approaches:** Leveson's System Theoretic Accident Methods and Processes (STAMP) [53] approach builds upon sociotechnical systems theory, expanding the scope of influences on systems out to the political level. The system is seen as having several hierarchical levels similar to Rasmussen's approach, each with its own control structure with controls and constraints operating vertically between the levels. Hazardous incidents are therefore seen as control failures. Though arguably better suited for the analysis of control failures in technical systems, the approach can also be used to analyse dependencies between system layers as understood in this report, for example inadequate control actions at the governance level leading to inadequate safety management practices.
- **Resilience engineering:** Resilience engineering addresses the issue that it is not only necessary to design systems in a way to reduce failure and prevent incidents and accidents, but also to ensure systems can function as required under both expected and unexpected conditions [54]. Safety-II is an approach that encourages an additional focus on ensuring that as many things as possible go right as a complement to the classical Safety-I focus of ensuring as few things as possible go wrong [55]. This approach is widely referred to as resilience engineering and it is of great relevance to the management of complex systems given its focus on managing uncertainty and emergent behaviour. It is important to note, however, that Safety-II concepts are not a replacement for Safety-I, but the philosophies should be considered in combination for completeness.
- **Function and performance variability:** The Functional Resonance Analysis Method (FRAM) [56] introduces a notion of functional resonance to complement causal models or theories. Functional resonance can be seen as (not necessarily intended) interactions or dependencies between functions; it moves away from a time-sequence model of causality. The approach is based upon the premise that normal variability in task performance can lead to unexpected and undesired consequences based on the dynamic nature of interactions within the system, in combination with performance variations of tasks that are coupled via input/output, control, constraint, resource and temporal relationships. This type of analysis may inspire approaches that could be integrated into our framework to detect and analyse subtle interdependencies between seemingly independent components of a system.

# 4.4 Application to complex systems

There are a number of factors that characterise complex systems and the framework (see Sections 2 and 3) that limit the suitability of the above techniques for analysing and managing the resulting risk of systemic failures. These can be summarised as follows:

- **Definition of the system of interest and system boundaries:** As described in Section 2, one of the challenges of complex systems is defining the scope of the system under consideration. This is often a pre-requisite and first step for many of the safety analysis techniques described above. The results of the safety analyses will therefore only be valid from the particular perspective of the chosen system boundary. Methods need to cater for uncertainties regarding system boundary definition or where the assumed boundaries are explicitly stated so that they can be considered in further analysis (for example when comparing competing concepts of the system scope). A closely related issue is the difficulty in managing the risk of unforeseen (ad hoc) systems that may spontaneously occur due to unplanned interactions or interdependencies between different components and their environment. In such cases, a perspective must be taken based on the enclosing environment to determine appropriate mechanisms to react to and manage risk in the presence of uncertainties and emergent behaviour (for example through effective regulation). Aspects of cognitive systems engineering techniques may encourage a more open approach to considering system boundaries and deriving measures at management and governance layers.
- **Semantic gap:** It is often not possible to specify the correct or desired behaviour of complex systems, either due to the complexity of the task itself or due to competing sets of objectives that lead to ambiguous notions of safe behaviour (also known as the semantic gap, see [1] for more details). The consequence of this semantic gap can be a lack of targets for acceptable levels of risk and uncertainty regarding what should be considered a failure.

  Due to the uncertainties related to the system boundaries and the specification of the intended behaviour of the system itself, there may not be a clear set of system objectives against which the risk can be assessed. Most safety analysis techniques, in particular FMEA, FTA, BTDs and STAMP, require on a set of system objectives against which the system can be measured. Due to the semantic gap, analysis approaches are required that identify and potentially continuously evaluate the appropriateness of emergent system objectives to ensure that they are in line with a set of broader expectations. Resilience engineering perspectives may help to ensure safe system behaviour despite uncertainty in the definition of underlying safety requirements.

  Inevitably this leads to the need for proactive and integrated regulatory approaches that can adapt and coordinate across industry sectors as system boundaries and expectations shift over time. Cognitive systems engineering in combination with the functional resonance method could potentially be adapted to support these approaches.
- **Unknown and unforeseen faults:** Many of the safety analysis techniques described above require some model of the system and in particular of how faults propagate to failures. However, one of the consequences of complexity is the presence of unknown and unknowable faults and causes of failures as well as a high level of inter-connectivity and non-linear interactions. Causal approaches to safety analysis and management, which at most can be used to increase system robustness, must therefore be complemented with approaches to increasing system resilience to allow for a reaction to unforeseen events and emergent properties. Safety II approaches of resilience engineering could potentially be adapted to address the unpredictability of emergent properties of complex systems.
- **Performance limitations and uncertainties:** Fault–error–failure-based methods such as FTA and FMEA are not well suited to analysing the effect of performance limitations (where no specific fault is present) in the system or where a sufficiently detailed model of the system does exist to allow an analysis of the propagation of faults and failures. The ability to model the impact of uncertainties, both in the environment as well as the internal behaviour of the system, is important [57]. This could lead to measures for uncertainty prevention, removal, tolerance and forecasting, for example, by analogy with the work by Laprie et al [45].
- **Changes to the system over time and dynamic behaviour:** The methods described above also typically do not consider changes in the system objectives, functionality, structure and environment over time. Such changes can arise from the impact of memory, path dependency, inertia, coupled feedback, mode transitions and tipping points in the system. An extended set of methods will therefore be required in order to assess the impact of these properties during safety analysis and will need to include a continuous approach to

evaluating the risk inherent in the system at any point in time. To better determine the impact of dynamic effects of the system on the risk of systemic failures, it will be necessary to determine appropriate observation points to measure leading indicators of critical changes in system state.

Factors that characterise complex systems limit the suitability of traditional safety engineering and management techniques. An extended set of techniques must therefore be developed and used in combination in order to counteract the increasing risk of systemic failures.

# 4.5 Guidelines for method selection

This report does not attempt to produce a definitive list of the most appropriate methods as this is not a practical solution to the problem of system complexity. Instead, we identify the need for an approach for selecting appropriate *combinations* of methods for application at and across the governance, management and task and technical layers. When doing so, it is important to understand the scenarios in which a particular method is applicable, the value each method brings, its pitfalls and how techniques can be combined to the greatest effect.

Key to this approach are the principles inherent in the development of a safety case and the implementation of a safety management system (SMS). A safety case provides a rational argument as to why a technical system is acceptably safe to operate based upon evidence. It is important, and good practice, as part of a safety case, to present the rationale for why certain techniques were sufficient to provide a credible set of evidence to underpin the validity of the argument.

An SMS describes a set of requirements and processes by which an organisation can manage the safety of its operations. While not structured as an argument, this organisational approach to achieving and assuring safety equally relies on a structured framework for selecting and applying a suitable combination of methods (albeit with a different scope to the safety argument) to deliver safe outcomes.

In both the construction of a safety case and the definition of a SMS, an overarching safety analysis approach and management strategy should be defined which, for a given set of risk criteria, selects a set of methods best suited to analysing and managing this risk. This requires a strong reliance on the use of suitably qualified people with the competency and experience to make judgments (under uncertainty) of the appropriateness of processes and techniques to support safe outcomes. These people will require a key set of skills that include system-specific knowledge, independent thinking and the ability to master techniques that promote creativity in order to identify underlying issues in the safety concept (such as false assumptions).

Ensuring that the principles inherent in safety cases and SMSs are applied at the governance, management and task and technical layers is a first critical step toward managing complexity. Appropriate method selection that can be justified by competent personnel is a fundamental part of achieving safe outcomes. However, it is also important to understand the limits of the current set of safety and risk analysis and management methods. Furthermore, applying purely analytical arguments will not be sufficient to communicate concepts of risk in such a way that those making critical decisions (for example in government) fully understand the consequences of those decisions. Neither will such techniques be sufficient to communicate risk concepts to those most likely to be affected by systemic failures but whose behaviour may have a significant impact on the overall risk. Therefore, careful consideration is required to take all stakeholders perspectives into account and to communicate the dependencies of risks within complex systems in an accessible manner. It will be important to consider the working in terms of uncertainty and trust, rather than risk, as means of widening communications.

We envision that the set of methods will include aspects of Rasmussen's layered approach in combination with a complex systems engineering-based understanding of the causal factors leading to systemic failures both within and across the layers. The framework presented here also has a strong relation to BTDs [21] in terms of highlighting the relationships between possible causes (of system complexity), their consequences and systemic failures, as well as the design and operation time controls to reduce the likelihood of causes or to limit their consequences.

The case studies that informed this report suggested that the interactions between the layers are a crucial component in triggering systemic failures. This can be seen as complementary to Perrow's *Normal Accident Theory* [58], which promoted the concept that systemic failures in complex systems are inevitable and mainly caused by management and organisational factors. This is illustrated in Figure 10, although it is not suggested here that further work should be limited to this viewpoint. It should be noted that these relationships are not necessarily linear or sequential but iterative and developing over time. Therefore, depending on the focus of analysis on the nature of the interactions within the system, alternative representations to those suggested in Figure 10 will be required to fully understand the interactions between the layers.

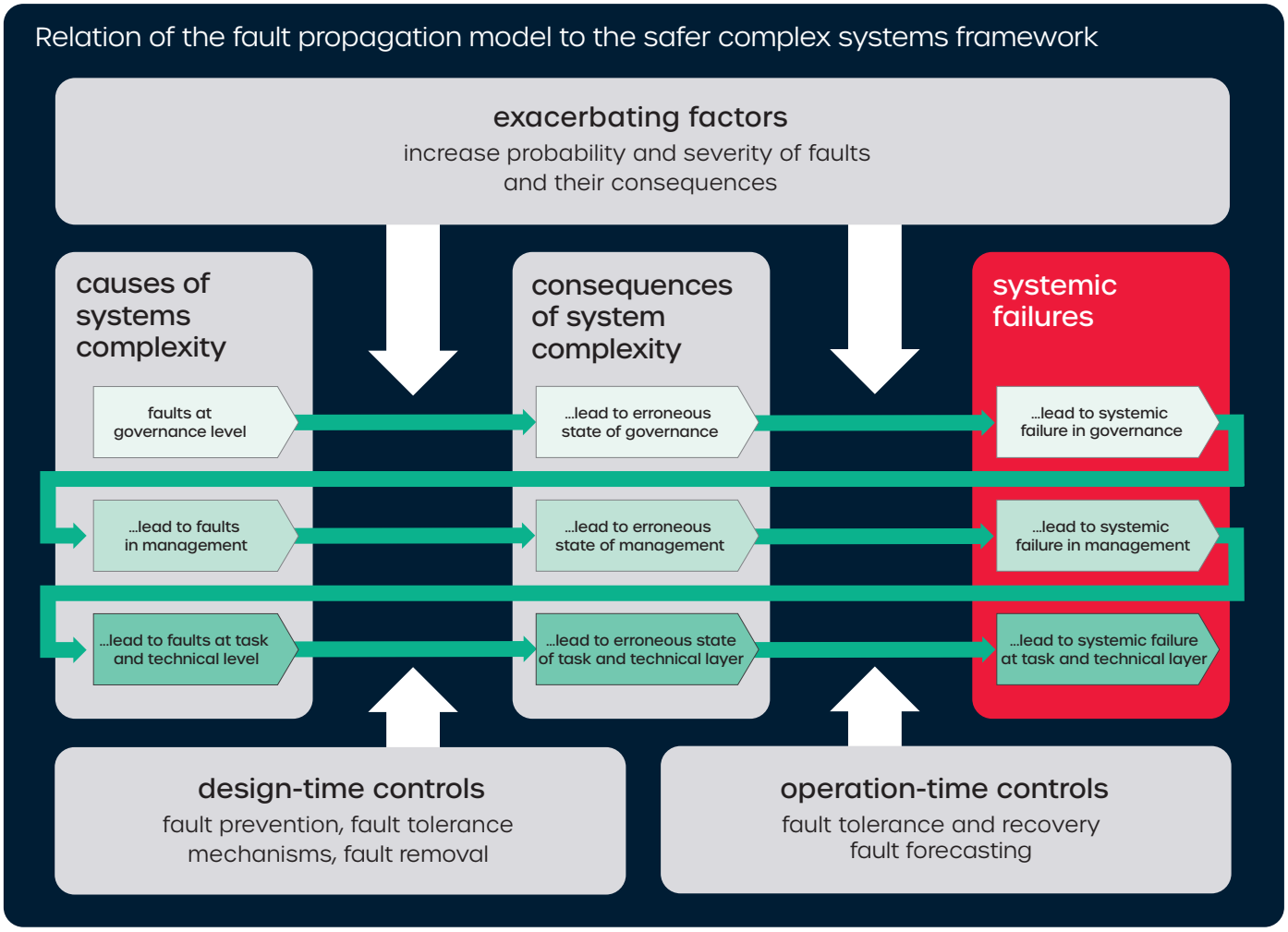



Figure 10: Relation of the fault propagation model to the Safer Complex Systems framework

# 5

# Sector-specific analysis

This section summarises characteristics of the industry sectors that were examined as part of this study. Each sector is evaluated in terms of major trends and drivers of complexity. A number of case studies from each sector also informed the work in this report, several of which were analysed based on the framework presented in Section 3. These can be found in Appendix C, along with a broader list of case studies across a wider set of industry sectors.

# 5.1 Aerospace

The global aviation industry plays a significant role in both the economic prosperity and interconnectivity of the world. Globally, safety is held as an extremely high priority and it is recognised that a safe aviation system contributes to the economic development of states and their industries. Safety regulation is coordinated through the International Civil Aviation Organisation (ICAO) [59] which also oversees the aviation system internationally, regionally and within states, with a strong focus on the prioritisation and continuous improvement of aviation safety.

At the heart of the aviation industry's safety performance is a strong focus on risk management in design and operation, supported by safety performance measurement. In 2017, aviation achieved the safest year ever on record [60]. However, in 2018 scheduled commercial air transport accidents resulted in 514 fatalities representing a significant increase from the figure of 50 achieved in 2017. Thus, safety improvement continues to be a focus.

As well as operational safety risks themselves (such as controlled flight into terrain and loss of control in-flight), ICAO lists both organisational challenges (including ensuring effective safety oversight) and appropriate infrastructure to support safe operation (as defined in the Basic Building Blocks Framework and enhanced by the Aviation System Block Upgrades) as key challenges that are under constant review by the international aviation community. Complexity is inherent within the risks that the aviation industry manages, the organisational structures that govern and manage the industry and the technical systems that form the infrastructure on which the industry operates.

There are both positive and negative lessons to be learned from how the aviation industry has managed the ever increasing complexity of the aviation system and continued to achieve a long-term and sustained improvement in safety performance (see Section 6.2). The effort required to maintain this improvement is significant given the rapidly increasing complexity within the aviation system.

### 5.1.1 Drivers of complexity in aviation

The global aviation system is highly complex. It exhibits most of the characteristics of complexity described earlier in this report, and often major aircraft accidents are an emergent result of the system and its interaction with the environment (recent examples include QF32, MH370, MH17, 737 MAX, see Appendix C).

Some of the key drivers of complexity in aviation include: highly safety-critical activities, the global nature of the industry; many human based activities and interactions; large complicated technical systems; highly interconnected processes; strong reliance by other industries; and increasingly rapid changes in technology. Many of the systems and processes that have assisted in creating the current safety performance are starting to come into conflict with technology advancements that do not fit the mould of how the industry has advanced historically. Technology advancement and disruption within (such as increasing use of autonomy) or in industries adjacent to (for example retail sale of toy drones) the aviation system have started to challenge the historically successful approaches to governing, managing, developing and operating its components.

The ability of the aviation industry to respond to the causes and exacerbating factors of complexity will be a key consideration in maintaining or improving upon the current safety performance as historical practices will be insufficient.

# 5.2 Connected and automated vehicles

### 5.2.1 An industry under change

The automotive industry is currently undergoing a period of radical change. The rise of mobility services coupled with the trend away from individual vehicle ownership is being driven by new business models, internet-based eco-systems and a diversification of vehicle types including e-scooters, automated package delivery vehicles and urban shuttles. These systems are enabled by a much higher level of connectivity to both the end-user and to traffic infrastructure and other road users. Mobility is no longer defined in terms of individual vehicles but instead by complex systems of systems. Multi-modal forms of transport, infrastructure-supported automation and smart motorways (see Appendix C) are all examples of systems of systems within an overall transport network. The resulting increase in interconnectivity as well as the dynamic and permeable nature of systems boundaries are key factors in increased complexity in the mobility sector.

Many of these innovations are driven not only by the need for more convenience and the positive environmental impacts of reduced traffic congestion: they also have a potential for greatly increasing overall road safety. Human error can be attributed to more than 90% of accidents on the road [47] and automated driving systems (ADS) have the potential for making roads significantly safer by restricting the impact of potentially inattentive and unreliable human drivers. However, there are several challenges in realising the full safety benefit of automated driving. An Ethics Commission [46] established by the German Ministry of Transport and Digital Infrastructure described the need to demonstrate a positive risk balance of automated driving technologies compared to average human performance. The Ethics Commission also called for a proactive driving style and the avoidance of discrimination based on person-related characteristics. Automated driving will improve performance in most situations but will not be able to eliminate the risk of accidents altogether. However, the judgement on whether or not automated vehicles are considered acceptably safe will not only be made through comparisons to human drivers. If it can be argued that applying state-of-the-art development approaches could result in a significantly better performance than an average human driver, then achieving current accident rates will not be an acceptable safety target for automated vehicles.

### 5.2.2 Drivers of complexity

In 2017, when Volvo [61] first started testing its ADS in Australia for the first time, it encountered something the Swedish designers had not necessarily anticipated – kangaroos. Having trained the system to accurately recognise and predict the path of mammals such as deer and elk crossing the road ahead, the movements of the marsupials responsible for 90% of the animal-vehicle collisions in Australia had the system stumped. Since then, there have been other incidents [62] of automated driving vehicles misinterpreting their surroundings with fatal consequences for the vehicle occupants and pedestrians. The requirement to accurately perceive a continuously evolving environment is just one of the drivers for complexity in automated and connected mobility solutions. These systems will also need to remain safe in the presence of unpredictable interactions with human operators and other road users as well as during *ad hoc* collaborations with other traffic systems and infrastructure. In addition, the increased use of artificial intelligence (AI) for perception and decision functions, especially machine learning (ML) introduces an additional element of irreducible complexity and uncertainty within the system (see Section 3.4.3). Public trust in such systems will also require them to act in a predictable manner and adhere to local traffic laws and conventions that, at times, may not only be ambiguously stated but may even need to be violated in order to reach a minimal risk state (which is how human drivers actually drive).

### 5.2.3 Legislation and standardisation

Current legal requirements do not adequately support automated driving scenarios, from the lack of machine-readable traffic laws to the unclear definition of liability in accidents caused by or involving automated driving functions. The European Technical Committee on Motor Vehicles [63] has defined guidelines for exemption procedures for the type approval of automated driving systems. In the USA, a state-by-state approach has been applied in approving the test and deployment of such systems on public roads. The National Highway Traffic Safety Administration has published guidance for the safety of ADS [64]. This includes a description of 12 safety elements that should be considered when using ADS on public roadways. System developers are encouraged to produce a Voluntary Safety Self-Assessment (VSSA) document that demonstrates how they have addressed each of the safety elements. The UN Economic Commission for Europe (UN ECE) has also published guidance on the regulation of automated driving functions, including detailed requirements on the test and type approval of automated lane-keeping systems [65].

Several organisations, such as ISO [66] and UL [67] are also developing technical safety standards for automated driving systems, leading to a partially competing, incomplete and fragmented set of guidelines that are unlikely to mature until after the first generation of products are on the road.

# 5.3 Healthcare

### 5.3.1 A system under pressure

Healthcare is the organised provision of medical care to individuals or a community intended to enhance quality of life. A report by three of the UK's National Academies on engineering better care [68] sees healthcare as: *"a set of elements: people, processes, information, organisations and services, as well as software, hardware and other systems that, when combined, have qualities that are not present in any of the elements themselves"*. Thus, healthcare can be seen, in the terms used in this report, as both a system and a system of systems. It includes primary, secondary and tertiary care in hospitals, but also treatments in the community. It is also influenced by many other systems, such as social security, government policies for controlling misuse of alcohol and drugs, and so on. While some of our concerns are very general, our emphasis is mainly on healthcare provision through hospitals.

Providing safe healthcare is challenging. No two patients are identical. Different diseases may present similar symptoms, making them hard to diagnose. Patients can have comorbidities, for example the presence of one or more additional conditions occurring at the same time as a primary condition, which make diagnosis and treatment more difficult. A patient's condition can change very rapidly. The notion of risk introduced in this report is still applicable, but often it is necessary to assess the risk of different courses of action – noting that doing nothing carries risk as patients' conditions can worsen, perhaps fatally, without treatment (see Appendix C.3.1 for an example). Further, in healthcare the focus is normally on safety risks to individuals unlike many other domains where the focus is risk to the population that can be adversely affected. Thus, although healthcare has guidelines and procedures for treating patients, such as clinical pathways, considerable care and judgement – including about risk – is needed in treating individual patients.

The availability of resource can be critical. Healthcare is expensive, with the Office for National Statistics (ONS) putting the costs in the UK at around 10% of Gross Domestic Product (GDP) and noting significant variations among other OECD Nations [69]. A World Bank analysis [70] shows huge disparities in healthcare spending as a proportion of GDP (2016 data). The USA is an outlier at 17.7% but rich nations average about 12.6% whereas sub-Saharan Africa is around 5.2% and Southeast Asia 3.6%; these differences are further amplified by differences in GDP *per capita*. There are other complicating factors, for example relative cost of labour in different parts of the world, and differences in distribution of spend, for example between capital and labour. Nonetheless, it is clear that there is a huge resource discrepancy between different nations and that developed world solutions may simply be unaffordable in other parts of the world, which means that providing healthcare solutions has to be cognisant of regional and economic factors. These differences in wealth and other factors such as living conditions have a significant impact on the provision of care and the prospects of patients in different areas of the world. Although many charities are seeking to address these issues, there remains a great disparity in provision between regions and income groups.

Even in the richer nations resources are finite. In the UK, the National Institute for Health and Care Excellence (NICE), *inter alia*, assesses cost-effectiveness of drugs and makes recommendations on which should be made available through the National Health Service (NHS). Further, in hospitals there is limited availability of: beds, intensive care unit (ICU) facilities, clinicians, medication, and so on. In general, decisions may have to be made about which patients to treat (including triage in emergency situations) and, on rare occasions, deciding who receives potentially life-saving treatment and who does not; this is clearly a risk-based decision. In the developed world, healthcare depends increasingly on technology. Surgical instruments are a long-established technology but, for example, recent advances have enabled minimally invasive surgery. In recent decades devices such as syringe pumps, which automatically deliver pre-defined doses of medicines or fluids, have become commonplace. There are now remotely operated surgical instruments, with the prospect of robotic surgery. Health information technology (HIT) such as electronic patient records (EPR) are used to replace paper-based systems, with the aim of reducing opportunities for human error.

Healthcare is also subject to ethical constraints. For example, clinicians must make decisions about palliative care and, perhaps together with family members, may have to decide when to stop use of life-support systems. There are also more controversial issues such as euthanasia with differences in legal frameworks in different countries and other legal constraints on healthcare provision.

Patient safety is a high priority in healthcare. In recent years, there have been initiatives to learn from practices in other domains such as aerospace, and to adopt and adapt established safety engineering practices in healthcare. For example, the US Federal Drug Administration (FDA) has produced regulations requiring the use of safety (assurance) cases for medical devices [71]. The aforementioned work on a systems view of healthcare safety [68] identifies many long-established safety

methods (from Safety-I) that could be applied in healthcare. Hollnagel and colleagues have placed a significant emphasis on Safety-II in healthcare as a way of achieving resilience [72, 73] and these ideas have been very influential. However, he is now acknowledging the need to treat Safety-I and Safety-II as complementary, a view endorsed in this report: see the healthcare recommendations in Appendix D.3.

Healthcare also has the concept of 'never events' that is "patient safety incidents that are wholly preventable" [74]. In the UK these events are itemised, see [75], and include, for example, "surgery at the wrong site" and "administration of medication via the wrong route". The NHS policy [74] is for 'never events' to be analysed to enable learning from experience (see also Section 7.1.4). It should be noted that analysing 'never events' is necessary but not sufficient to enable learning from experience. It is also necessary to learn from near-miss events. The discussion of the UK's NRLS [76] in Section 6.2 shows the difficulties of collecting such data in healthcare. Further, so far as can be ascertained, there is no requirement to predict the frequency of 'never events' – this can be viewed as healthcare being 'reactive' rather than estimating risk in a way that is done in other domains. Estimating risk might require clinicians to address uncomfortable questions such as allowable patient fatality rates – somewhat analogous to setting safety targets in other domains, for example aerospace. Analysis of 'never events' and near misses are valuable in learning from experience – but it seems hard to engender a 'learning culture' in healthcare. The establishment of such a culture seems crucial to improving patient safety and an explicit recommendation is introduced in Appendix D.3 along with a recommendation on data analysis.

Healthcare can be viewed both as a system of systems and a sociotechnical system, with significant ethical and resource constraints – it is under pressure anyway, but the COVID-19 pandemic raises this to an unprecedented level and on a global scale. The two case studies presented in this report (see Appendices C.3.1 and C.3.2) illustrate the challenges on very different scales. The first involves a single patient with sepsis, a condition where diagnosis and treatment is difficult and time critical and, in this case, complicated by ethical and legal issues. This emphasises the fact that healthcare has to explicitly consider risk to individuals, in contrast to other domains, and also highlights the need to consider the risks of doing nothing. The second considers the COVID-19 pandemic, written about four months after the initial cases were identified in Wuhan, China, when the virus had spread to almost every country in the world. The report's aim is not to enter the realm of epidemiology, but to highlight some of the systems aspects; a sub-problem related to the provision of PPE is also presented in Section 3.3 as an illustration of the framework. The following sections consider drivers of complexity, some of which provide context to these two case studies.

### 5.3.2 Drivers of complexity

There are many drivers of complexity in healthcare. The discussion here cannot be comprehensive; instead, it aims to illustrate the wide range of factors driving complexity and some of the coping strategies.

First, there are many factors that make diagnosis and treatment of illness complex. These include diet, climate, living conditions, and the nature of the healthcare system. For example, the spread of disease is likely to be different in very crowded areas such as the favelas in Rio de Janeiro, compared to rural areas with low population densities such as Saskatchewan in Canada, and affluent areas in major cities. Many nations have healthcare systems provided (largely) by the state but in others, notably the USA, the healthcare system is largely private, accessed via insurance, and this introduces disparities based on the ability to pay, and thus obtain insurance (President Obama's initiatives notwithstanding). These, as well as individual economic circumstances, are all aspects of equality, diversity and inclusion. Second, although medical science is very advanced, in some respects, there are many things the medical community does not know, and the problems we are addressing change over time, for example as viruses mutate. In some areas there is an adoption of a systems perspective for example, rather than viewing comorbidities as merely complicating factors; some research seeks to categorise them [77] in order to systematise their analysis to help in diagnosis and in defining treatment strategies. In the context of this study, this shows an attempt to address (and provide a control for) one of the drivers of complexity.

Third, clinical decision-making has to be made in real time against changes in a patient's condition, such as sepsis, where speed of response is critically important to the clinical outcome. Also, when there is limited availability of treatment facilities or medication (resource) clinicians must undertake triage to decide which patients to treat. On a broader scale, the transition of a disease to become a pandemic can be seen as a tipping point when the speed at which the disease spreads potentially exceeds the ability of the healthcare system to manage it.

Fourth, artificial intelligence (AI) is now used in healthcare, for example, in some image analysis tasks to facilitate diagnosis [78]. There are AI-based systems,

such as the so-called AI Clinician [34], that automate aspects of prescription – in this case for vasopressors and intravenous (IV) fluids used in the treatment of septic shock.

At present, such systems are at a pre-clinical stage and it is not yet clear how to assess safety sufficiently to allow them to be deployed in hospitals, nor how to balance the AI system's recommendations against clinical judgement. However, this is an active research area which is likely to contribute to the complexity of healthcare in the near future (although it might also constitute a form of control).

Fifth, healthcare has an increasing focus on evidence, with the term evidence-based medicine (EBM) [79] being used for the treatment of individual patients, whereas evidence-based healthcare has a broader interpretation, focusing more on evidence at the population level. For example, data from randomised control trials (RCT) are used to assess the effectiveness of drugs, and RCTs are often viewed as the gold standard for effectiveness evaluation [80]. Further, data generated from healthcare systems, such as HIT, can be used to support modelling and predictions, for example, of epidemics to predict spread and the impact of different treatment strategies. This can be viewed as a control rather than a driver of complexity, but it does contribute to complexity of the healthcare system as a whole.

Finally, there are different systems of medicine and one can contrast traditional Chinese medicine (TCM) with Western medicine. TCM has its origins more than 2,000 years ago and places an emphasis on the whole body 身体 (she¯ ntˇı) and the ability of the body to heal itself, assisted by natural treatment such as herbs. In contrast, Western medicine focuses more on the treatment of individual diseases, and draws heavily on science, for example in the development and evaluation of drugs. This is not the place for an extensive comparison of the two systems, but some say that TCM is better for chronic conditions and Western medicine is better for acute conditions, where rapid intervention is needed. At a minimum, there are different views of what constitutes the system and different views of the evidence base. Despite these differences, there is some debate about the benefit that could be obtained from combining these two systems of medicine.

Healthcare can be viewed both as a system of systems and a sociotechnical system, with significant ethical and resource constraints – it is under pressure anyway, but the COVID-19 pandemic raises this to an unprecedented level and on a global scale.

# 5.4 Supply networks: food, water, power, and money

Supply chains are not linear, therefore supply chains are often now discussed more in terms of *complex networks* rather than chains, recognising complex (sometimes bi-directional) interdependencies between organisations that are difficult to definitively map [81]. The money supply in the finance sector is one example of a poorly mapped complex supply network [82]. It is also worth noting that even vertically *coordinated* supply chains, like the supply of some food items where there is a clear path between producer and consumer, that coordinated chain exists within a *complex network* of suppliers that provide support services (such as parts, labour and other resources).

Safety of supply networks as complex systems manifests itself in a number of different ways. Two significant aspects are the safety and integrity of the supply network itself (which overlaps significantly with what is often termed supply network/chain *resilience*), and the safety and integrity of the materials circulating through a supply network. The case studies covered in Appendix C.4 represent more of an analysis of the second of these aspects, the integrity of the materials flowing through a supply chain. However, our ability (or lack of it) to map the complex interdependencies of supply in a supply network is a contributing factor in both case studies. It is also of significance to our understanding of the resilience of a supply network, as it is hard to determine how exposed a supply network is to failure if you do not have a good understanding of all the dependencies at each *node* in the network, since complex networks may not fail in predictable ways, and may demonstrate behaviours such as cascading failures [83].

Safety of supply networks is also affected by a diverse range of other factors, including, but not limited to: problems with ageing infrastructure that is difficult or expensive to replace (such as power supply networks); and changes in management and loss of institutional memory, which could manifest in a lack of accountability of management and/or unclear responsibilities.

### 5.4.1 Systemic shocks to supply networks

The lack of a full understanding of supply network dependencies, and potential exposure to failure, also has consequences for supply network safety during systemic shocks (such as COVID-19 [84]). Systemic shocks impact the whole, or large parts, of the supply system at all levels, rather than individual parts. Systemic shocks therefore also frequently affect the economy and other systems more broadly, which can have unpredictable effects on supply networks. This highlights the fact that supply networks are *open systems*, and as such do not really have clearly definable boundaries, for example where one supply network ends and another begins, or where the supply network ends and the wider economy or society begins.

Systemic shocks can therefore expose problems and cause safety issues in supply networks, that would under normal conditions not be apparent. For example, *lean* supply networks (which remains in one form or another a dominant paradigm for supply chain management) are broadly intended to be frictionless (or close to frictionless), and therefore highly dependent on the flow of materials and resources. Very little redundant resource is kept at different points in the supply network (although there can be redundant supply in some cases). During a systemic shock it is possible that large parts of the supply network will be disrupted, hence the flow of resources will be adversely affected. A lean supply chain could therefore experience something like resource deadlocking, where the lack of redundant resources means that parts of the network stop as they are waiting on resources from elsewhere. The complexity of the interdependencies between the parts of the supply chain mean that releasing that deadlock could be very difficult as there is no obvious or single control that can be applied [85].

The other safety challenge presented by systemic shocks to supply networks (which is linked to the above) is the desire, or necessity, to substitute components in a supply network for other technologies. These substitutions could be less understood, newer/untested, of lower quality, and so on. However, the trade-off is to either stop the supply and wait for the original technology to be available (and risk resource deadlock and failure), or substitute the technology to allow the supply chain (the flow of resources) to continue but with a potentially inferior technology. This represents a competing objective that is not straightforward to solve. Furthermore, this competing objective operates across levels. At the local (organisational) level there is the desire to keep that organisation's production going, without compromising safety, which is a local management issue. There is also a systemic objective, to keep the entire supply chain safe and functioning, which is a *distributed* management problem. This would normally be managed by regulation of, for example, component safety standards; however, during a systemic shock how these competing objectives are managed is very difficult and would require coordination. One example of this is the substitution of N95 face masks with KN95 masks for PPE, which is discussed in Appendix C.3.2 and in Section 3.3.

Substitution of technology is not just a concern during systemic shocks: care must be taken whenever a technology is swapped or

reused. This has been a problem in aerospace, with some substitute parts that do not meet the high standards for aviation use. As a further example, technologies are frequently reused in the software industry, where tools and methodologies can be taken from one domain and used in another. If this is done then there needs to be awareness of the underlying assumptions around the use of that technology, and whether the new use case will violate any of the assumptions made in the initial domain of use.

### 5.4.2 Drivers of complexity in supply networks

Complex networks have the capacity to exhibit all the features of complex systems, discussed and defined in Section 2 and in Appendix A.2. In particular (as discussed above), they increasingly exhibit the features of complex networks. In a complex supply network the organisations are analogous to the parts of a system, and the interdependencies are the relationships between those parts. Therefore, supply networks will, for example, have emergent properties that cannot be predicted from only an understanding of their relationships with organisations. Their behaviour will also be non-linear, and therefore how they respond to shocks (perturbations) cannot easily be predicted, as has been seen with COVID-19 (see Appendix C.3.2 and also [86]).

Supply networks are deeply embedded in societies, and therefore are reflective of many aspects of the culture and values of that society. A driver of complexity that only increases as supply networks cross between societies, potentially coming into conflict with other cultural and societal norms that they are less compatible with or not sensitive to. This can be an additional source of competing objectives.

The multiple jurisdictions that international supply networks cross present a problem for developing any standard practices, methods, or regulatory bodies that can provide oversight of complex supply networks. This also extends to agreements around how and where complex supply networks are monitored. Enforcing a standard of monitoring of a supply network when all or part of it is outside your jurisdiction is very difficult, although there is work on using modern technologies, such as blockchain, on ensuring provenance and integrity in complex supply networks.

Supply networks are complicated by multiple competing objectives. Significant to safety is the *lean* operation of supply networks, where efficiency can be traded against resilience, or supply traded against quality/safety. This could reduce their resilience to things like systemic shocks, by reducing redundancies in the supply network that could allow it to continue operating through periods of resource disruption.

Uncertainty in supply chains is also a systemic factor that contributes to their complexity. Uncertainty makes it difficult to have complete oversight of all the input into any one product, as the sourcing would have to be known at each point (node) in the network. That sourcing may also change in times of resource scarcity, or simply for economic reasons, when parts of the supply network might substitute suppliers and/or technologies.

### 5.4.3 Infrastructures for developing an understanding of safety of supply networks

Our understanding of supply networks, and therefore the safety of supply networks, is to a degree dependent on having sufficient data available for them, along with the tools, methods and theory available to effectively use that data. The data itself presents a problem: as yet there is no consensus on who should collect what data and in what format, and the international nature of supply networks would make developing a consensus on any of those points challenging. Modelling and simulation tools and methodologies could, and should, be developed to use what data is available to understand how supply networks behave and might respond to systemic shocks. Tools that would support and go beyond the systems engineering and other tools currently used to model and control supply networks [87]. However, even if this is possible, determining who is responsible for acting on the insights from any modelling or simulation will be difficult. As again, networks cross multiple jurisdictions (sectorial, political, regulatory), and decisions taken in one jurisdiction will produce effects elsewhere in the system. There may be benefit in looking at other sectors, such as the finance sector and stress testing, for possible solutions (or routes to solutions) to these problems. However, this broadly remains an area where significant work needs to be done, and the security of our supply networks requires it is done.

The report discusses supply networks in more detail in section C.4, with two examples of *E. Coli* C.4.1 and PFAS Chemicals C.4.2. Recommendations specific to supply networks can be found in section D.4.

# 6

# Findings

This section distils the results from the case studies, stakeholder engagements and wider research to identify a set of main findings and define a vision for the future direction of the **Safer Complex Systems** programme. Where possible, the evidence for these findings is identified by referring to other sections of this report or by providing external references. The text in ***italics*** refers to elements of the framework presented in Section 3.

# 6.1 Finding one: Public safety relies on increasingly complex systems

**Finding one: Public safety relies on increasingly complex systems**

Systems that have an impact on public safety are growing in number, complexity and interdependency. Influences at the governance layer include growth in systems that are deployed and/or can operate globally,

For example telecommunications systems such as Loon balloons, which operate in the stratosphere and thus span *multiple jurisdictions*. At the management layer, supply networks are becoming more global and dynamic, including using technologies from one domain in another, exemplifying *supply networks and cross-domain collaboration*. At the task and technical layer, complexity is also being driven by increased *interconnectivity and interdependence* and *disruptive technology (AI and autonomy)*. All of these tend to increase complexity and are also occurring in a context of growing societal expectations of safety.

Systems are growing in complexity across many domains. While it is possible to provide examples at all three layers in the framework, it is easiest to illustrate this at the task and technical layer, particularly emphasising the technological perspective. Within the automotive industry, the design of electronic control systems is currently undergoing a revolutionary change. The functional pressures of increased electrification, automation and connectivity [88] are resulting in a shift away from distributed towards centralised system architectures; and towards much more sophisticated hardware platforms and software architectures that had not previously been qualified for use in safety-critical embedded applications. In simpler terms, over the last 10 years, the typical software size in an average passenger vehicle has increased from around 10 million to over 100 million lines of code, with another 10-fold increase predicted in the next decade. This software will continuously evolve over time and be updated via over-the-air mechanisms, often without the knowledge of the end-customer (indeed, over-the-air updates are already happening on high-end vehicles). At the same time, this shift is also leading to a fragmentation and reorganisation of the supply networks and partnerships involved in developing the systems. This in turn is leading to new business models and allowing startups and incumbents from other domains to enter the market.

Software size is not the only measure of complexity, but it is one that can be seen in many sectors. For example, in aerospace there has been a growth in software size between successive generations of aircraft – both military and civil, as illustrated in Figure 11. In terms of sheer software volume, cars are now more complex than aircraft, although the data in Figure 11 does not include in-flight entertainment systems. However, it seems unarguable that the rate of change in the automotive industry is now far outstripping that in aerospace.

While not all domains are seeing similar growths in complexity, the impact of the pervasive underlying technological advances including telecommunications, cyberphysical systems (CPS), AI/ML (artificial intelligence/machine learning) and the Internet of Things (IoT) is felt in many sectors internationally. The rate of growth of connected IoT devices illustrates this, as shown in Figure 12.

IoT and other communications-based technologies are significant, not only because they enable growth in complexity of individual systems but because they enable interconnections that make otherwise independent systems interdependent. Such *ad hoc* or unplanned systems often have no clear ownership or allocation of responsibilities and accountability, such as the complex ecosystem surrounding 'white goods' (see Appendix C.10), which causes difficulties at the management and governance layers.

The rise of *ad hoc* or *accidental* systems, in turn, gives rise to unanticipated emergent properties. In such systems the dependencies may only become apparent after serious consequences are discovered. Examples of such effects can be seen in several of the case studies, for example Lancaster power outages [89] and the GM ignition switch problem [90] (see Appendix C).

## Growth of software complexity in aerospace systems

Thousands of source lines of code (KSLOC) used in specific aircraft over time

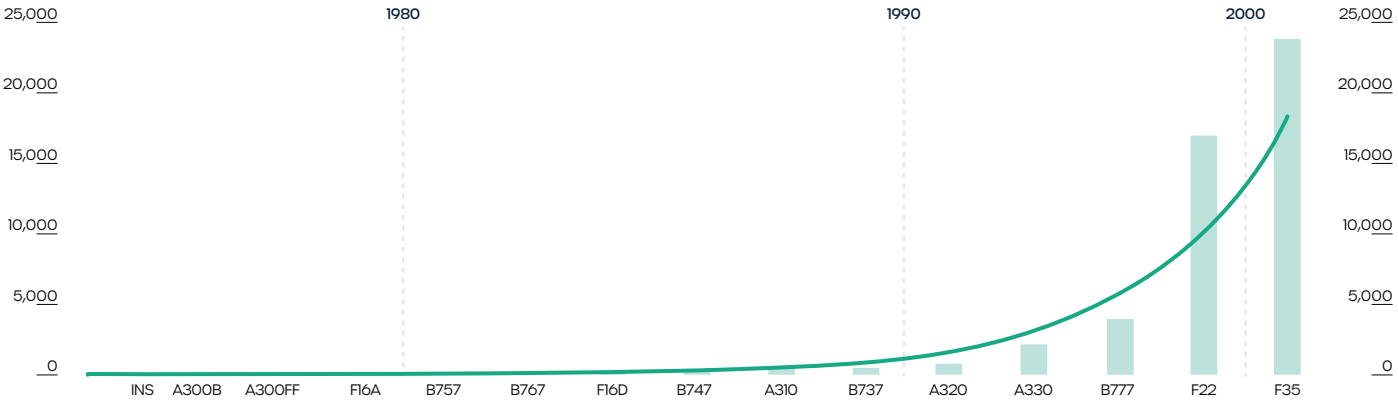



Figure 11: Growth in aircraft software size

Political imperatives can have unanticipated impacts across unconnected systems. For example, decarbonisation and the push to net zero emissions can lead to changes in use of different transport modalities, and hence impact safety (perhaps for the good), but also introduce new challenges. There is a growing demand on the electricity infrastructure for vehicle charging. However, many prefer to charge overnight when electricity charges are lower and, as this is a time when solar energy is unavailable, the system is less able to cope with the demand.

Complexity gives rise to (implicit) *risk transference* between users/ stakeholders, which can lead to *accountability and moral responsibility gaps*, for example between the developers and operators of built infrastructure. Also, interdependencies can expose limitations in the regulatory regime. For example, prompted by the COVID-19 outbreak the International Labour Organisation (ILO) has identified 'regulatory gaps' related to the "prevention of diseases caused by biological hazards" [91].

In the health sector, the regulatory framework is very complex – even just in the UK. A recent report arising from a 'regulatory sandbox' on 'machine learning in diagnostic services' identified 13 different bodies with regulatory roles, including the Care Quality Commission (CQC), the Medical and Healthcare products Regulatory Agency (MHRA), standards bodies, and the Information Commissioner's Office (ICO), in relation to the use of data. While this is just one domain, similar issues are seen in other domains and the interplay of regulatory responsibilities illustrates one of the governance challenges in ensuring safety of complex systems.

## Internet of things - active connections worldwide 2015-2025

Internet of Things (IoT) active device connections installed base worldwide from 2015 to 2025* (in billions)

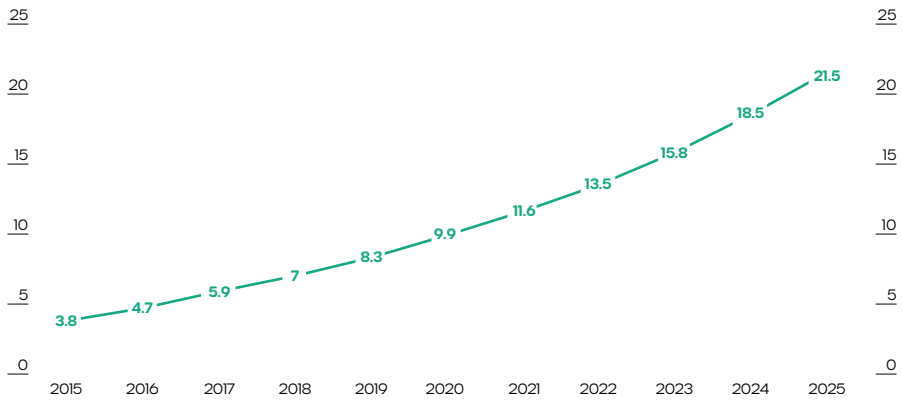



Figure 12: Growth in number of connected IoT devices over time

## 6.2 Finding two: Historically, some domains have seen sustained improvements in safety

**Finding two: Historically, some domains have seen sustained improvements in safety**

In some mature domains, systems have been remarkably safe and have shown sustained improvements in safety over many years, despite their growing complexity. However, this has not been true in all domains and this also depends on how we scope the *system of interest*. Further, as seen in Finding 4, the growth in complexity means that changes are needed to ensure this remains true in the future. As well as providing evidence for this assertion of 'historic safety', it is important to seek to understand why systems are so safe in order to define a future direction for the **Safer Complex Systems** programme. Some initial observations are made here and a more systematic analysis is presented in Section 6.3.

Accident data are collated and published across several industries. The following examples show the downward trends in accidents in a number of sectors, but also some of the variations internationally and the difficulties of data collection. Boeing publish accident statistics annually, which include data on the worldwide civil aircraft fleet. The latest data covers the period 1959 to 2018 [92]. The report contains many graphs but Figure 13 best illustrates the global trends in both accidents and fatalities (not all accidents give rise to fatalities).

There are several reasons for the decreases in accident rate, but one of the primary factors is the thorough analysis done following any accident and the drive to learn lessons, not just for the particular aircraft type or airline but across the industry, thereby *learning from experience*. Indeed, the aviation industry is often cited as the paradigm for accident and incident investigation, and this is an example of the value that can be obtained from a Safety-I mindset. While the approaches used in this industry cannot be applied without change in other industries (see the discussion of healthcare below) it can be seen as an example of good practice to inform the development of reporting and analysis systems in other domains. However, if the *system of interest* is expanded beyond aircraft to include the environment, then it can be said that there is a negative impact on safety through contribution to global warming; this is the type of indirect effect alluded to in Section 2.2.

Another area where improvements in safety can be seen is in road transport. The number of fatalities on the roads has been steadily declining for decades. According to the Department for Transport (DfT) there was a total of 1,770 road deaths in the year ending June 2018. This represented a decrease of 35% over a 10-year period. In addition, as can be seen in Figure 14, the overall number of accidents experienced by drivers has also significantly decreased. More than 90% [47] of traffic accidents can be attributed to human error. However, the positive trend showed by the statistics is unlikely to be based purely on improved driving skills. Around 77% of accidents (based on DfT statistics) occur because of the driver's inattentiveness or recklessness. Therefore, it can be assumed that other factors such as improved traffic infrastructure and technical driver assistance systems contributed both to the reduction in the number accidents and their severity. This has happened with the initial introduction of so-called 'Smart Motorways' in the UK [93], where personal injury accidents were shown to be reduced by more than half during the trial period (see also the discussion in Appendix C). Further, public attitudes to safety and the introduction of the European New Car Assessment Programme (EuroNCAP) [94], pressure from insurers and other influences are likely to have contributed to these long-term safety improvements. However, as the number of automated vehicles on the roads increases, future accident statistics will require a different interpretation in order to gauge the effectiveness of safety measures and to derive appropriate responses. Also, as with aircraft, if one takes a broader view of the *system of interest* then there are negative safety consequences of road vehicles through the impact on air quality [95]. Finally, it should be noted that the 'drivers' of long-term improvements in safety in the aerospace and automotive sectors are quite different so care should be taken in trying to 'translate' experience from one domain into another.

Although there is a positive trend in road fatalities in many countries, a significant variation is seen internationally.

The data for healthcare in the UK is very different. Figure 15 comes from the National Reporting and Learning System (NRLS) [76] and shows data from a voluntary incident reporting system that has been in use since late 2003 and is slowly growing in usage. Thus, the figure should not be viewed as showing a worsening in outcomes but a growing acceptance of the value in reporting incidents and accidents. However it is in stark comparison to the reduction in accident rates seen in aviation, which can be attributed, at least in part, to effective accident investigations.

The NRLS report [76] also contains other useful summary data,

## Accident rates and onboard fatalities per one million departures

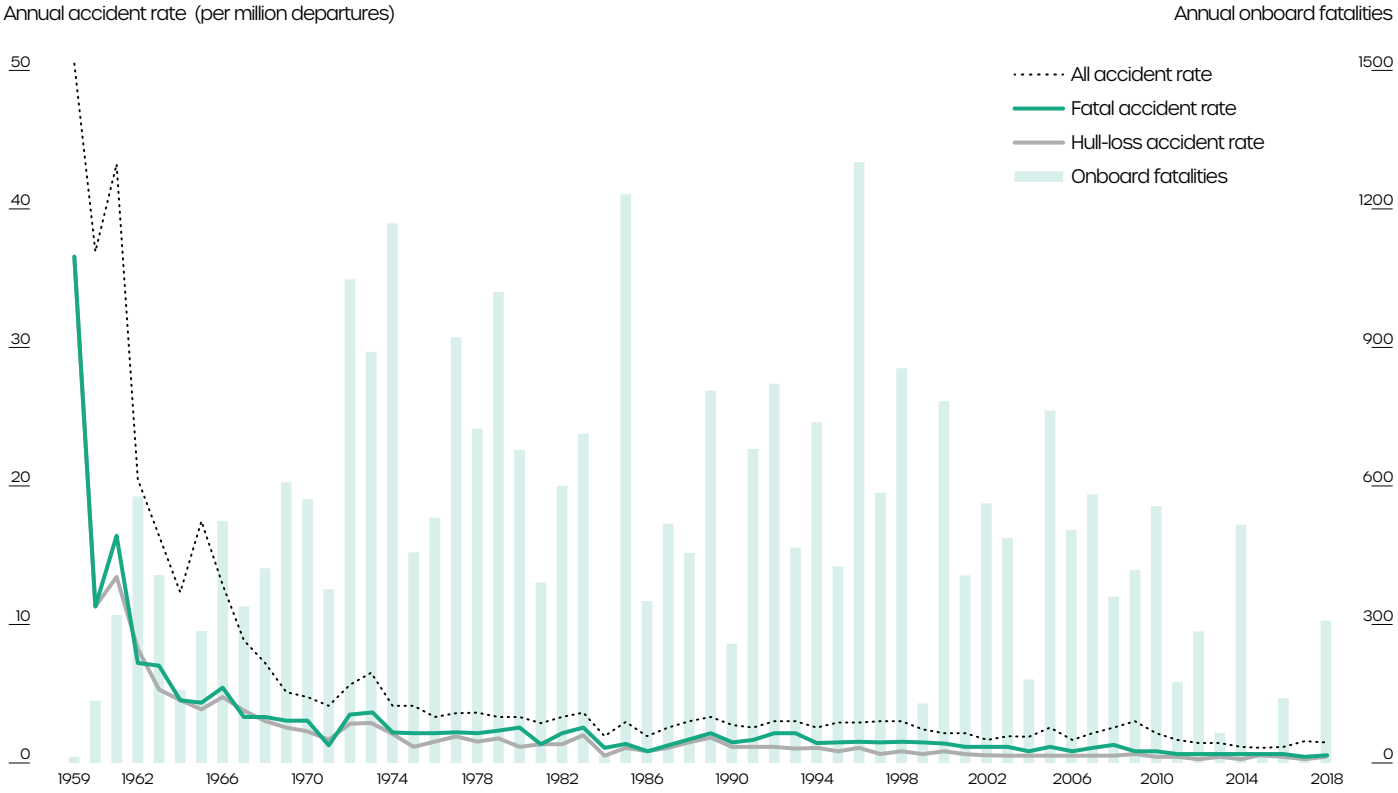



Figure 13: Global Air Accident Trends

including the distribution of incidents by severity and type, for example in the period October 2018 to September 2019 about 10.6% of the 1.9 million incidents were medication errors.

The figures above show that, in mature industries, accident and fatality rates are generally dropping, albeit the rate of decrease is slowing in some sectors. This is not to deny the significance of catastrophic events such as the 737 MAX accidents and the fire at Grenfell Tower. It might be that these are 'outliers' against a downward trend in accidents. Alternatively it might be that they indicate 'tipping points' that signify the need for urgent action because of weaknesses in regulation, as is occurring following the Grenfell Tower fire with the establishment of a new Building Safety Regulator (BSR). The NHS data is included to show that it takes time to put in place incident reporting systems and care should be taken when interpreting data during the early stages of introducing such systems.

**Global inequalities in road accidents**

Although the number of road accidents and deaths has been declining for decades, the global reduction in road fatalities has slowed since 2013. This could be caused by increased traffic related to economic progress, reduced law enforcement efforts and increased popularity of cycling. However, there are huge disparities in the progress on road safety between countries. For example, in the period 2010 to 2017 Norway saw a 50% reduction in road fatalities and Greece 42%, while 90% of global road fatalities occurred in low- and middle-income countries [96]. This figure is disproportionate relative to the countries' level of motorisation as they account for only 54% of the world's registered motor vehicles; the risk of fatal accidents in African countries is almost three times higher than in Europe [97].

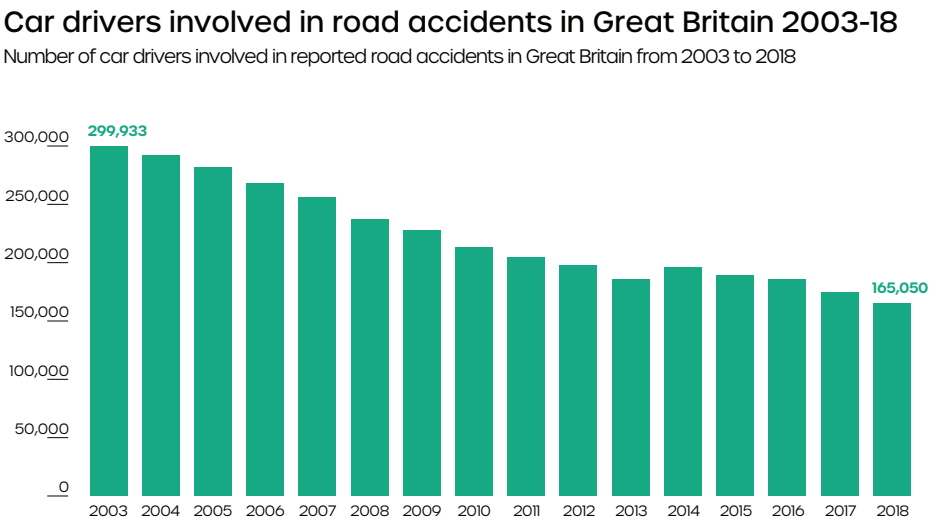



Figure 14: UK road accident trends, source: UK Government Department for Transport

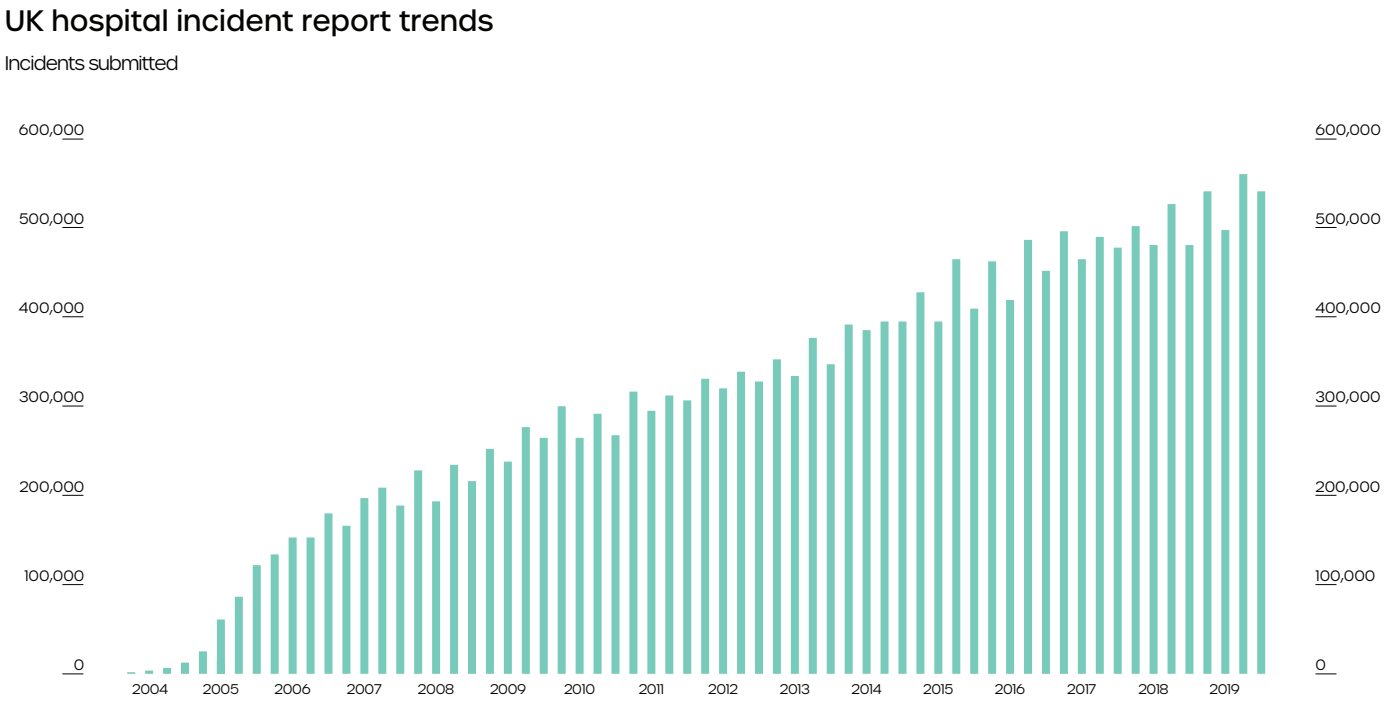



Figure 15: UK Hospital Incident Report Trends

# 6.3 Finding three: Safety management controls exist and are effective where they are used

**Finding three: Safety management controls exist and are effective where they are used**

It has previously been observed that systems are "remarkably safe" despite the limitations in safety analysis methods and standards [98]. The key factors cited [98] span engineering (*task and technical*), *design-time controls* (for managing system complexity), *management,* and operation-time controls. The stakeholder engagement particularly highlighted the importance of *operation-time controls* for managing complexity safely. It was generally accepted that this was necessary as the design-time activities were never sufficient (see Section 6.4). For most complex systems, there are additional issues that can only be addressed in operation and operational controls may be the only option for *ad hoc* systems. Those factors are included here, mainly under Operations 6.3.3.

The structure in [98] does not align exactly with the framework used in this report, but it can be mapped readily and the key observations from [98] are covered in the following sub-sections. They are then considered again in Section 6.4 to show how they are impacted by complexity.

### 6.3.1 Design-time controls

The factors referred to as 'engineering' in [98] map mainly to the design-time controls in the framework and include:

- **Domain knowledge** – domain knowledge is a key factor in identifying and removing requirements errors, including safety-significant ones.
- **Evolving products** – evolving successful designs is a way to reduce the problems of identifying hazards and hence controlling safety risk.
- **System architecture** – the more critical the system, the greater the importance of architectural defences, especially redundancy (sometimes known as fault tolerance, see Sections 4 and 6.3.3) in achieving safety.
- **Safety in systems engineering** – seeing safety as an effective tool in systems engineering, guiding the design.
- **Conservatism** – not using 'leading edge' technologies, to avoid the uncertainties of novel designs.
- **Control of engineering maturity** – achieved using technology readiness levels (TRLs), together with design for manufacture and prioritisation of problems based on their safety criticality.

Some of these 'design heuristics' are challenged by growth in complexity (see Section 6.4).

### 6.3.2 Managerial controls

The 'management' factors from [98] cover both design-time and operation-time controls, and are mainly at management layer in the framework:

- **Priority of engineering** – 'doing the right thing' almost regardless of other constraints.
- **Good leadership** – giving appropriate priority to safety in terms of resources, listening to concerns, and so on.
- **Supportive/just culture** – listening to the concerns of engineers and responding to problems constructively, regardless of level and status.
- **Developing and rewarding competence** – ensuring that individuals with appropriate skills are recruited and encouraged to develop professionally.
- **Impact of regulation** – knowing that systems are independently and effectively regulated helps to 'keep the organisation honest'.

Although the latter point is presented from a managerial perspective it clearly relates to the governance layer. It was also noted by some of the stakeholders that regulatory activity appeared to have more impact than standards.

### 6.3.3 Operation-time controls

The factors referred to as 'operations' in [98] are *operation-time controls* in the framework introduced here:

- **Fault tolerant operation** – at (almost) all times an element of the system has failed or is operating below full capacity/capability (in the case of humans) and the system operates successfully despite this.
- **Good operators with good training** – operators know how to deal with the system in normal and failure modes.
- **Empowerment** – operators have the authority to make difficult decisions, including suspending operations.
- **Time** – time between initiating events and accidents enables operators to assess the situation and to plan and implement appropriate remedial actions.
- **Simple mitigations** – despite system complexity, many hazardous failures have simple remedial actions, for example switching off.
- **Learning from experience** – reporting operational issues and removing/mitigating both the immediate and root causes (see Section 6.2).

Note that some of these resonate with aspects of equality, diversity and inclusion (see Section 3.4.2). Further, these operational controls are generally viewed as increasing the resilience of the system.

# 6.4 Finding four: Increasing complexity threatens existing management controls and governance capabilities

**Finding four: Increasing complexity threatens existing management controls and governance capabilities**

The growth in complexity challenges many of the reasons why some systems 'are so safe' (see Section 6.3). Growing complexity of designed systems and the emergent complexity of *ad hoc* systems are outstripping our engineering methods and challenging our ability to manage systems safely. A range of issues, including the exacerbating factors in the framework, are bringing us to a *tipping point*. These drivers of complexity should be a cause for concern – good historic safety trends may no longer be achievable. These concerns are discussed by considering which of the controls described in Section 6.3 may cease to be effective as systems increase in complexity. The concerns are sufficient to suggest the need for sustained and focused activities to develop additional controls, across all three layers of the framework, so systems will continue to meet societal expectations of safety.

### 6.4.1 Design-time controls

All controls are more difficult to implement as systems become more complex, but the following are particularly affected by the drivers of complexity.

- **Evolving products** – *rapid technological change* makes it impossible to evolve products and harder to identify hazards and assess safety risks.
- **System architecture** – with some modern technologies, for example AI and ML, it is unclear how to make classical architectures such as duplex and triplex work (can we learn different 'models' that are close enough in behaviour so we can compare their results?), although it may be possible to use functional redundancy.
- **Conservatism** – there are many novel designs using 'leading edge' technologies, such as AI and ML, which are hard to assure because of the *weak science base*.
- **Control of engineering maturity** – classical approaches such as TRLs don't work with some novel technologies and there is a move away from classical lifecycle models on which such controls are based. For *ad hoc* systems there is no obvious way to use such measures.

These design-time issues place even greater emphasis on managerial and operational controls.

### 6.4.2 Managerial controls

Several of the managerial controls identified above, such as priority of engineering, are applicable to complex systems, but many are challenged by growth in complexity.

- **Good leadership** – as systems become more complex, and more novel, it is harder to provide good leadership as experience regarding appropriate resource levels, the right mix of analysis methods to use and so on cease to apply (see also Section 4.5).
- **Developing and rewarding competence** – this is still a valid control, but much harder to achieve as there is a growing *competency gap* as the skills needed are in short supply and generally not part of normal professional education.
- **Impact of regulation** – there are now significant *standards and regulations lag* because it takes time to develop new standards and regulations, especially where international consensus is needed, and currently growth in complexity is outstripping the ability of the affected industries to respond.

These are all problematic, but the gaps in regulatory frameworks are perhaps the most far-reaching in terms of their impact (see Section 6.4.4).

### 6.4.3 Operation-time controls

As noted above, the challenges of complexity further emphasise the importance of *operation-time controls* – but these too are affected by complexity.

- **Empowerment** – in some cases, such as with autonomy, there is an impact on *human oversight*, for example the operators may not have the situational awareness to make effective decisions and autonomy may deny operators the authority they need to implement corrective actions, as seems to have been the case with the 737 MAX (see Appendix C.1.2).
- **Simple mitigations** – the interdependencies between systems and the difficulty of understanding the source of emergent properties make simple remedial actions much more difficult, if not impossible, to identify, for example with 'Forever' chemicals (see Appendix C.4.2).
- **Learning from experience** – generally past experience is much less relevant as systems are both novel and changing – in the worst case previously learnt mitigations may be inappropriate or impracticable. Again the 737 MAX appears to illustrate this problem (see Appendix C.1.2).

Perhaps the issues with *learning from experience* are the most concerning, as it undermines a

strategy of evolving successful organisational practices.

### 6.4.4 Safety engineering

This study is concerned with safety of complex systems – it is thus instructive to consider how complexity affects our ability to conduct safety analysis and safety assessments.

First, quantitative approaches to evaluating risk based on statistics, in particular pre-deployment, lose credibility (if they haven't already [38]) and it will be increasingly necessary to evaluate risk more frequently, in some domains continuously during operation, as systems or their environments evolve and learn.

Second, at present there are no adequate processes and methods for analysis and managing the safety of complex systems, particularly those involving disruptive technological innovations. This is true for designed systems, and systems of systems, but particularly so for *ad hoc* collaborations between systems. The normal tactic of using appropriate combinations of methods (see Section 4.5) is still appropriate but it is less clear what combinations to use and if the methods available address all the factors that arise with complex systems.

Third, because of the greater connectivity of systems, for example using 5G networks and the IoT or CPS, there is a need to consider cybersecurity and its impact on safety. Methods have been developed over a period of time [99], [100], for addressing the interaction of cybersecurity and safety but these are not yet widely adopted in many domains.

Fourth, the safety methods do not cope well with change. The cost of repeating safety analysis after change is high (often comparable with initial analysis costs) and this has led to attempts to achieve incremental certification [101] but that has proven difficult even with today's systems. With complex systems, change is almost continuous, whether this is because of replacement/repair of faulty components, training of operators, or upgrades to technology or functionality. It is far from clear how this problem can be addressed, although there is a growing research emphasis on dynamic risk assessment [102].

Fifth, techniques for operational safety management of complex systems are relatively well-established, and principles for resilience management, disaster recovery and so on can be migrated from mature industries to those that are less mature, at least in some cases (see also the recommendations in Section 7). However, there is limited ability to identify leading indicators to warn of impending problems, and hence to take early action. Often, even where these indicators do exist, bias in risk-perception or political imperatives hinder an adequate or consistent response (as can be seen with COVID-19 where there have been widely different responses from different governments, see Appendix C.3.2).

Sixth, the legislative and regulatory frameworks for safety are largely inadequate for dealing with the growing complexity of systems. The regulations do not deal well with the new technologies and may even militate against using them. In many cases the evolving and interconnected nature of the systems means they span regulatory boundaries, thus involving multiple regulators, see for example [103]. This suggests that it may be necessary to revise regulatory structures, potentially combining regulators so that there are fewer regulatory bodies, but each has a wider span of responsibility.

Finally, there are limited skills and experience to deal with the demands of complex systems across all layers of the framework from developers and safety engineers through to management and regulators. There is also a problem where, for example, experienced safety engineers are unfamiliar with technologies such as AI and ML but the developers of such technologies have little understanding of safety (I or II). Solving such problems not only requires education and training but also ensuring equality, diversity and inclusion in teams working on such systems, and perhaps 'reverse mentoring' where junior staff mentor more experienced staff who are unfamiliar with the emerging technology.

### 6.4.5 Safety culture

The safety culture of an organisation can be thought of as the product of individual and group values, attitudes, competencies, and patterns of behaviour that determine the commitment to, and the style and proficiency of, an organisation's health and safety management. Pragmatically, culture is very important in achieving safety, as the case studies show – both positive examples such as NATS (see Appendix C.1.1) and negative such as 737 MAX (see Appendix C.1.2). Culture will become even more important as complexity increases as it will not be possible to fully preplan and rehearse treatment of safety issues. Therefore, the way in which the organisation responds to unanticipated, emergent properties will be crucial. This applies at all layers in the framework – in governance as much as management and task and technical, although the ideas are most mature in terms of operational organisations.

The concept of *high reliability organisations* (HRO) is very relevant to safety culture, and is described in terms relevant to system operations. More specifically, HRO

is often described in terms of five principles:

- **Preoccupation with failure** – continually focusing on "what could go wrong" and updating designs and plans as necessary.
- **Reluctance to simplify** – for example, treat problems as opportunities to learn and seek out 'root causes' and do not be satisfied with finding proximate causes.
- **Sensitivity to operations** – continuously monitoring operations to identify any potential deviations from objectives, including safety.
- **Commitment to resiliency** – as noted above, it is not that errors or failures never occur, but that the organisation is not debilitated by the events.
- **Deference to expertise** – listening to those with most knowledge, noting that they may be quite junior, as they are the ones with the 'hands-on' experience.

Note that there is a significant overlap with resilience and these can be viewed as those characteristics of organisations that help to ensure operational resilience.

Equality, diversity and inclusion have a key role here, for example in *deference to expertise*, as do concepts such as just culture [104]. For example, the culture should be supportive of all staff reporting issues related to emergent properties and avoiding responses such as "that couldn't happen" or "that doesn't happen to me" where different groups are affected differently by the system. A further issue will be a culture of considering "others that might be affected by my behaviour" and being aware of potential interdependencies with other systems (this is somewhat redolent of the 'general duties' arising from Section 3 of the Health and Safety at Work Act (HSWA) [105]).

Wide embedding of current good practice in developing and sustaining safety culture will be important for future complex systems. However, there is also a need to better integrate an understanding of equality, diversity and inclusion and to broaden the horizons of organisations in terms of who might be impacted by the system. This is strongly linked to the understanding of risk, where people need to think of risks to others, not just to themselves.

# 6.5 Observations

The growing complexity of systems means that it is becoming increasingly difficult to provide assurance of their safety. This implies that there are some classes of system that we should not develop and deploy, because they are beyond our ability to assure. However, it is hard to determine this boundary, except on a case-by-case basis. Worse, *ad hoc* systems might 'arise' that are beyond our ability to assure but as they are emergent, not designed or planned, then there is no regulatory 'control' to prevent deployment. Thus, there is a need for a 'mechanism' to recognise such situations and to act as a 'trigger' for relevant authorities to initiate remedial action, noting that there may be no relevant regulator if the emergent system crosses regulatory boundaries. In this case, it is likely that the emergent issues will have to be dealt with at a governmental level within national jurisdictions and through international collaboration where the system has international or global reach. In the UK, the most obvious recent example is the use of the Scientific Advisory Group for Emergencies (SAGE) to address the challenges arising from COVID-19. SAGE involves the Chief Scientific Advisors from across government departments, complemented by additional specialists, such as epidemiologists, and representatives of key bodies, for example Public Health England, and has collaborated with its international counterparts in addressing the management of COVID-19.

We need a mechanism (trigger) to recognise situations where *ad hoc* systems arise, unplanned and undesigned, so that the relevant authorities can initiate action to ensure and assure their safety.

# 7

# Recommendations

This section identifies a set of sector-independent themes seen as necessary to introduce a complex systems approach to thinking about safety management across the governance, management, and task and technical layers. It is suggested that the themes should be addressed through future phases of the Safer Complex Systems Progamme, and example grand challenges and research areas are outlined.

# 7.1 Sector-independent themes

This section identifies a set of themes seen as necessary to introduce a complex systems approach to thinking about safety management across the governance, management, and task and technical layers. We recommend that these themes should be addressed in more detail during the next phase of the **Safer Complex Systems** programme, both through dedicated activities to refine the topics described below and further case studies to collect more data and validate the overall framework. While being sector independent, the themes will require specialist competencies to refine them as part of future work and may best be addressed through a range of activities from basic research through to direct engagement at the policymaking and legislative level.

However, we acknowledge that each sector will have also have its own unique set of challenges and will require specific solutions based on sector-specific expertise and methodologies and the report includes several examples to better illustrate these issues. Appendix D contains a number of sector-specific recommendations and additional observations that refine the themes presented here. It is intended that these will form a basis for more focused work in each sector.

### 7.1.1 Theme one: Risk, trust and acceptable levels of safety

**Develop approaches for better communicating risk, increasing trust and forming consensus on acceptable levels of safety**

As systems become more complex, the concepts of risk and acceptable levels of safety become harder to define. There needs to be a greater emphasis on understanding and articulating acceptability of risk, particularly in relation to systemic failures that are, by their very nature, hard to predict. A common language for communicating risk is required that can be shared among policymakers, industry and laypeople in order to reach consensus for setting safety targets and to build trust in the systems and/or in the organisations that develop, operate, sustain, and regulate those systems.

Approaches to articulating risk must be developed that can be used not only by policymakers, regulators and safety professionals during design and operation of the systems but can also be used to engage the general public. In doing so, differences in risk perception [49] must be considered that may be because of cultural sensibilities or access to relevant information. This requires dialogue with all affected stakeholders and equality, diversity and inclusion has a key role ensuring that the views of all groups are considered and accommodated. Thus, involving representatives of key stakeholders will be important for future complex systems, but challenging where the scope and sphere of influence of the system is hard to bound.

There is a related issue of safety and risk awareness, that is the appreciation of the presence of risks, or that risks apply to particular groups of individuals. The COVID-19 pandemic has illustrated the effects of variances in risk awareness in the willingness to heed hygiene and social distancing advice. This includes how risk perception can vary over time and can be influenced by a number of factors such as xenophobia [106] and political influences. Although too early to draw conclusions, early results [107] indicate that more must be done to ensure that as large as possible a portion of the population understands and acts on the (medical as well as more wide-ranging economic and societal) risks associated with such events. Furthermore, the COVID-19 pandemic illustrates the links between apparently independent ecosystems, including an impact on the healthcare supply chain (see Sections 3.3 and 5.3), demonstrating the need for engaging with a broad range of stakeholders beyond what is commonly understood as the system scope – and again it may be more helpful to think in terms of the sphere of influence.

Risk management measures for complex sociotechnical systems can only be effective if there is sufficient trust in the effectiveness of these measures and in the judgement of those responsible for proposing these measures. A lack of trust in the system may itself become a significant contributor to risk. As an example of how this issue can manifest itself at the governance layer, see the public response to COVID-19 related restrictions and how lack of trust in either the science base or elected and non-elected officials' own personal actions can undermine the discipline with which such measures are applied or can even be enforced. At a task and technical layer, a lack of trust in the performance of an automated driving system coupled with a

**Sector-specific illustration – automated urban mobility**
A new set of standards and guidelines are currently under development to regulate the safety of automated vehicles, both on the road and in the air [63] [64]. There is currently debate within the standards community regarding which safety targets to set for such systems. An Ethics Commission [46] established by the German Ministry of Transport and Digital Infrastructure described the need to demonstrate a positive risk balance of automated driving technologies compared to average human performance. However, society is unlikely to accept such systems if they were to consistently collide with certain groups of people because of programming errors or inadequate training data [109], regardless of the statistical probability of the event occurring. As well as finding the right statistical basis for quantitative development and operation-time controls, there is a need to agree on a set of safe behaviours of the system that will also actively increase the trust of the vehicle passengers and other traffic participants alike. This is especially true for ambiguous situations that will require systematically evaluating diverse perspectives from legal, ethical and engineering perspectives [1].

lack of transparency regarding the system's control decisions can lead to a driver actively or unwittingly working against the system, leading to conflicting actions between the driver and the vehicle, which may lead to loss of control of the vehicle. In time, it might be that finding ways of articulating and communicating trust becomes more important to managing safety of complex systems than more traditional concepts of risk.

Applying a common method of communicating and discussing risk is a prerequisite to achieving consensus on acceptable levels of safety for complex systems and therefore also a determination of the actual system risk (see also Section 7.1.5) and an adequate set of control measures for achieving defined targets. One possible approach to achieving consensus is to apply reflective equilibrium [1], so as to reach/maintain a balance through negotiation between stakeholders.

As this process takes time it is perhaps most appropriate at the governance layer, although it may also have a role in management – especially if it is possible to do agile reflective equilibrium to help manage complex situations. The challenges in understanding and communicating risk can also be addressed in terms of uncertainty, including the use of the Johari window, see Figure 16.

The ability to communicate and inform stakeholders of system risk is also related to the topic of appropriate or calibrated trust [108]. Undue trust in the system can itself lead to risks as is evident in the phenomenon of automation complacency (see Appendix C.2.2) or the impact of misleading or mixed messaging regarding COVID-19 prevention measures. A human factors view is therefore not only critical in addressing risk communication issues but also in analysing the impact of trust within the system.

### 7.1.2 Theme two: Complexity in oversight, regulatory structures and policymaking

**Acknowledge and address complexity in oversight, regulatory structures, legal accountability and policymaking**

Government policymakers should consider the growth in complexity and the trends in the scope and capability of systems, when examining regulatory structures. This should include the application of outcome-based standards and publicly available specifications as a means of increasing agility in regulation. Furthermore, issues surrounding tort law and the allocation of accountability across multiple stakeholders, each of whose actions may contribute to harm caused by a systemic failure, should be addressed. In some cases, oversight of the systems may be ill-defined or distributed, particularly in the case of *ad hoc* or accidental systems that cross traditional boundaries. By regarding regulatory frameworks themselves as complex systems, an evaluation of the effectiveness and inherent risks of regulatory failures should be continuously performed in order to consider changes in the environment and emergent risks of new classes of both engineered and accidental systems.

Traditionally, safety and risk assessment are time-consuming processes; further standards and regulations evolve slowly. However, systems are now evolving quickly, and the use of IoT, CPS and modern communication technologies means that systems can change composition (open systems) or be formed in an *ad hoc* manner requiring far more rapid change in (formal) risk assessment than is done, or is possible, in current practice.

The lag between the introduction of disruptive technology and effective regulation, for example through standards and legal precedents, must therefore be narrowed. This will require increasing agility in regulation where there may be a need to change rules rapidly, for example to introduce a requirement for controls over particular emergent properties. There will be value in taking a more proactive approach, such as through application of horizon scanning by government policymakers and regulators to identify changes in systems and technologies that could be deployed to identify required adaptations to regulatory practices.

Some changes are already being seen. For example, there are regulatory sandpits exploring new ways of regulating systems, in areas such as healthcare [103]. And the British Standards Institution (BSI) is producing Publicly Available Specifications (PAS) [17] in much shorter timeframes than is the norm for standards. In general, a shift towards outcome-based regulation and standards to decouple the definition of the required level of safety and strength of argument from the (technology-specific) processes for achieving these claims should be pursued. More fundamentally, there may need to be change in the basis for regulation approaches themselves. For example, rather than regulating via safety cases, regulation might be carried out through approval of operational safety management systems, or by permissioning (giving an organisation approval to conduct specified operations), associated with a strong incident and accident reporting and analysis regime. This implies a move away from cause-effect style of control through regulation towards a more systematic view of emergent risks and appropriate controls. In evaluating and restructuring regulatory frameworks, the Johari window (see Figure 16) could be applied; for example by applying permissioning approaches where uncertainty is low and confidence in the behaviour is high (the Open quadrant in the Johari window), but proscription in the cases where there is a high level of uncertainty in the task to be performed by the system, as well as in understanding the system's behaviour (the Unknown quadrant). In contrast, assurance cases can be used in the Blind and Hidden quadrants.

Taking an even broader view, the shifting and diffuse boundaries between previously disjointed sectors may require a more radical restructuring of regulatory frameworks in order to address emerging classes of risk from these systems of systems. For example, multi-modal transportation might best be addressed by having a single (national) transport regulator, rather than a regulator for each mode. Furthermore, the application of safety engineering principles to the regulatory system itself should be used to identify underlying causes of risk of systemic failures at the governance layer and systematically identify approaches to increasing the effectiveness of regulation in an uncertain environment.

To establish liability for harm, civil law requires it to be shown on the balance of probabilities (greater than 50%) that a legal person performed an action that caused harm. This can make the allocation of liability and thus compensation for damages to victims difficult to achieve. The Uber Tempe crash, described in Appendix C.2.2, demonstrated the difficulties in apportioning blame where systemic failures were present across all layers (task and technical, management and governance). This resulting liability gap results from the inherent complexity of the system and its operating domain. This is compounded by the failure to treat an autonomous system

as a legal person for tort purposes [1]. The framework proposed in this study combined with a set of interdisciplinary analysis approaches across the layers (see Section 7.1.5) could be applied when performing *post hoc* analysis and allocation of blame and liability. More radically, there could be need for a change to the legal framework around liability – and this might be a topic to be addressed later in the programme or referred to a body such as the Law Commission in the UK.

Finally, there may need to be changes in regulatory structures, for example, producing a single national body covering all modes of transport, or reduction in the number of regulators involved in healthcare. Detailed recommendations on such forms of regulatory change are outside the scope of this report, but they are being considered by a Lloyd's Register Foundation foresight review on the future of regulatory systems. Once this review is published, its findings and recommendations should be reviewed to assess their relevance to this theme.

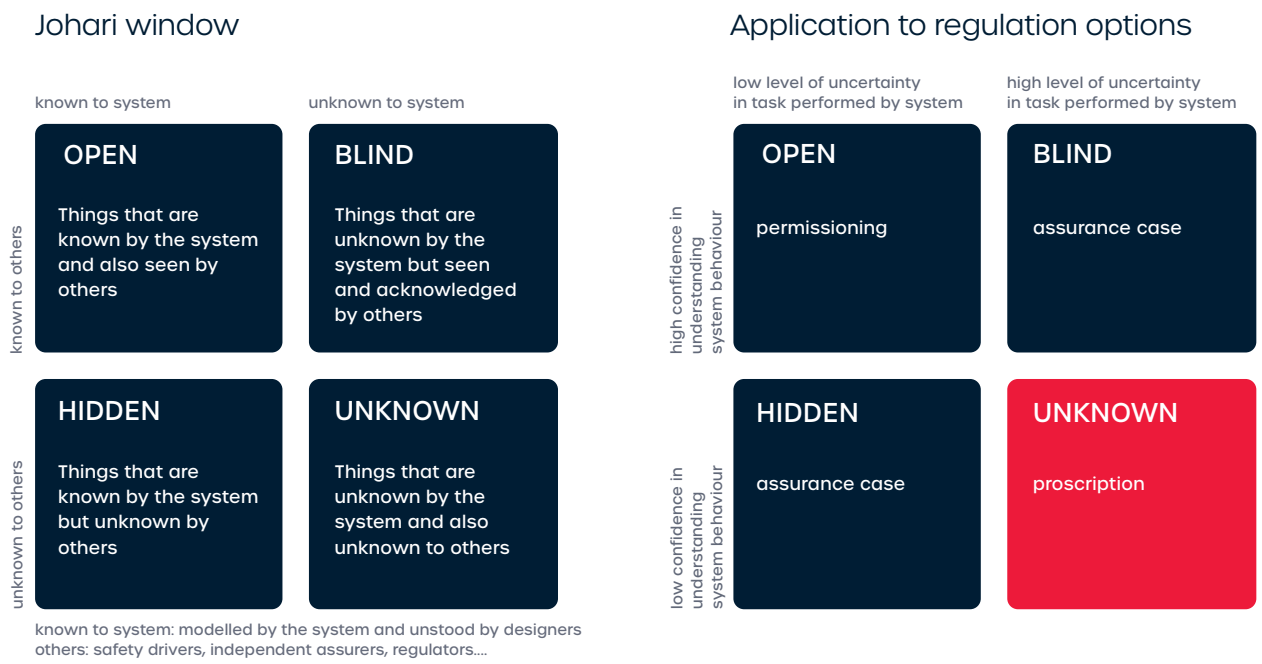



Figure 16: Johari Window and regulation mechanisms for complex systems

### 7.1.3 Theme three: Addressing equality, diversity and inclusion

**Develop methods to address equality, diversity and inclusion during risk management and promote heterogeneity of thought**

This report has highlighted the influence of equality, diversity and inclusion in several ways. Firstly, it has been shown that risk is not equally distributed between stakeholders in a system in relation to various diversity characteristics. Although often related to ethnic and economic background, gender or ability, some of these characteristics, or at least their relationship to risk, may not always be obvious because of the impact of system complexity. An explicit recognition of diversity and engagement with a wide range of perspectives during risk analysis and management is required, supported by an appropriate methodology. Furthermore, means should be developed to include heterogeneity of thought in risk management. This includes the recognition and support of whistleblowers but should go beyond this, involving a wider spectrum of stakeholders when formulating regulation or safety standards.

Our goal must be to make complex systems safer for every member of society. Approaches must therefore be developed across the governance, management, and task and technical layers to identify contributing factors that lead to disproportionate risk to specific groups in society associated with systemic failures, both in terms of the likelihood of risk exposure and the severity of its consequences. This report has highlighted several examples of the correlation between risk and diversity, most notably COVID-19, see Appendix C.3.2 and the box below. Furthermore, Section 3.4.2 has suggested some ways in which these issues could be addressed. Nonetheless, a more thorough and systematic evaluation of equality, diversity and inclusion in ensuring the safety of society would be of value in future work.

**Sector-specific illustration impact of COVID-19 on Black, Asian and Minority Ethnic (BAME) communities**

There is increasing attention within the press at the time of writing on the disproportionate impact of COVID-19 on various communities. This includes increased risk of exposure as well as lack of access to healthcare, for example in the favelas of Brazil and slums of India. However, in developed countries such as the UK and USA, there are clear signs of the BAME community being disproportionately impacted. While many of the factors contributing to this effect do not require a study in complexity science in order to be explained (for example access to healthcare plans and high proportion of BAME workers in essential jobs where social distancing is not possible), others are harder to explain. For example, in the UK there have been a disproportionate number of deaths of medical staff from the BAME community compared to white colleagues. According to government statistics, 20% of the NHS workforce are from a BAME background. However, an analysis by the Guardian newspaper, dated 16 April 2020 [110] found that 68% of the those NHS professionals who have died from COVID-19 were BAME.

One exacerbating factor that makes this analysis difficult is the lack of ethnicity-related data [110] recorded as part of official COVID-19 statistics (this is related to the discussion in Section 7.1.4). A systematic analysis of the impact of COVID-19 on minority groups as well as regional differences may profit from applying the principles of the framework outlined in this report to discover underlying causes and exacerbating factors.

### 7.1.4 Theme four: Data-driven prediction of systemic failures

**Integrate simulation, model-based analysis and digital twins into design and operational-time controls**

Data collection and analysis techniques need to be developed to enable the development of digital twins of complex systems that will allow for systemic failures to be predicted and their underlying causes to be analysed. This will involve applying a variety of techniques from model-based simulations to statistical analysis and machine learning. Such models should then be used to examine the effects of proposed changes to the system or its environment to predict limits of manageable safe behaviour. Mathematical modelling techniques to understand and predict systemic properties should be grounded in a practical understanding of how abstract properties can lead to systemic failures to meet specific system objectives; they should therefore be considered as part of an overall modelling and analysis strategy. The framework proposed within this study should be used to provide context for such a strategy.

One of the primary challenges of managing the safety of complex systems is to detect when a system is about to make the transition into a hazardous state (systemic failure). This is made difficult by non-linear effects between the inputs and outputs of the system, as well as hidden and unpredictable interconnections and interdependencies resulting in tipping points. Loosely coupled components can be more easily removed to help to return to stability. For highly coupled components, it is more difficult to restrain their feedback effects or disentangle them from the system. Sometimes we want to harness complexity and allow for a certain amount of chaos at the edge of manageable complexity. But how do we monitor whether we are reaching the limits of benign emergence? This requires sensing the right set of indicators and identifying patterns of 'normality'. Anomalies may be very subtle and time sensitive. In particular, we often don't know what 'typical' behaviour is in *ad hoc* or accidental systems (of systems). Some properties of the system relating to interconnectivity can be analysed using mathematical techniques during design, thus allowing for certain predictions to be made about the likelihood of complex interactions within the system leading to adverse effects. However, this requires some knowledge or estimation of system properties in the first place.

Most future complex systems will be data rich. In many industries, including aerospace, there is already extensive data collection and analysis. This is not just the black boxes used for accident analysis but systems collecting health data to diagnose failures and, to an extent, predict failures allowing pre-emptive maintenance activity, thus avoiding failures. Engine health monitoring (EHM) capabilities have been in development since the early 2000s [111] and now the systems are now becoming much more sophisticated leading to the IntelligentEngine concept [112] enabling much more active management of availability, as well as safety.

Analogous capabilities are starting to appear in other sectors, such as with building information modelling (BIM), which is a computer-based process that enables architects, engineers and operational professionals to more efficiently plan, design, construct, and manage buildings and infrastructure. It is key that the BIM models transition smoothly from design to operation, avoiding unintentional risk transference. A more general term, digital twins, is used for developing computer-based models to help in the design and operation of systems, for example to help in managing the health impact of pollution in cities [113].

There are also examples of using the data to help in safety performance measurement and management. The UK Health and Safety Executive (HSE) has a project, funded by Lloyd's Register Foundation, to "unlock the potential of health and safety data" [114]. Increasingly ML is used to find unrecognised patterns in operational data. For example, Bayesian network analysis has been used to expose the causes of patient safety incidents in post-operative care where patients have undergone thoracic surgery [115]. Also, work has been done using ML for risk prediction [116] that, interestingly, notes that the techniques work for known knowns and unknown knowns, but not for other aspects of uncertainty – presumably because even with unknown knowns it is clear what factors to use in the ML-based analysis. Most of these approaches are analysing data retrospectively – this is necessary but will not be sufficient for future complex systems.

Data-based analysis will also involve designing the systems to be measurable and predictable in the first place so that leading indicators can be monitored and evaluated. Ultimately the aim should be to enable *learning from experience* before, during and after system operations – as noted in the evaluation of the Haiti earthquake [117]. To learn *during* operations will require ongoing safety performance

measurement and an issue will be the extent to which analysis can be conducted in real-time. For learning *after* (and hence *before* other operations) there will be a need, as part of management and governance procedures, to define data to be retained and communicated from systems that will be capable of producing vast amounts of data to enable accident and incident investigation. This is being considered for autonomous vehicles [17] but needs to be done more widely.

Complex network modelling and simulation methods are also being applied in several sectors. For example, complex network methods have been applied to the analysis of accident data in construction projects [118], as a novel method to understand near-miss time series data. Complex networks approaches have also been applied to the problem of cascade failures in supply networks [119], with a cascade failure in the Italian power grid being an example [120, 121]. Complex network methodologies are being proposed as a new way to theorise resource supply networks more generally [81].

### 7.1.5 Theme five: Holistic approaches to risk assessment

**Develop an integrated and complementary set of methods for analysing risks in complex systems**

Deriving strategies and associated methods for analysing and managing safety risk associated with complex systems is obviously a key area of future work. This will involve extending existing approaches and looking beyond these for a set of complementary techniques that compensate for deficiencies or limitations of current methods. This could include techniques that cover what formal and technology-focused analysis can or cannot do. Techniques can include storytelling, rich pictures, simulation and reflective equilibrium (see, for example [1]) to ensure a diverse set of opinions from different stakeholders are captured when determining the risk associated with the system. Holistic approaches to risk assessment will require cross-disciplinary and cross-sector viewpoints, including from the technical, sociotechnical, human factors, economics, communication, system-theoretical, and mathematics domains.

This study has shown how the risks inherent within complex systems can originate across the different governance, management, and task and technical layers with typical systemic failures resulting from uncontrolled consequences of a combination of factors across these layers. A holistic view of risk is therefore required. There is also a need for greater agility in risk assessment, for example the ability to respond rapidly to events, especially those that were unanticipated, and to the realisation that an *ad hoc* system has been formed or has emerged – and thus needs to be understood and managed. However, at present, models used to assess risk are typically static, disjointed and focused within specific layers and domains. Current quantitative approaches lose credibility, in particular when applied before the system has entered into operation, while qualitative approaches are restricted because of both the lack of appropriate system models and the widespread desire for more quantitative assurances about risk.

Cross-disciplinary approaches are therefore required to integrate focused risk assessment techniques into a common framework to allow the various perspectives to inform each other, leading to a better overall understanding of risk. This will include an effort to categorise existing models and explain how and when they can be used in combination. For more details see Section 7.2 for proposed research in this direction, as well as Section 4 and Section 6.4.4.

Part of the future work to develop a set of appropriate methods should include developing a taxonomy of causes and consequences of system complexity, as well as the types of systemic failures that complexity can lead to. This taxonomy can then be used as a set of guidewords to help conduct such an analysis. Initial suggestions for these taxonomies can be found in Section 3, especially Section 3.5. These would then be applied in combination with domain- and layer-specific models for analysing specific properties of the system and for ensuring that diverse opinions, as well as the needs of all stakeholders, are adequately catered for. Of course, this also needs to consider the perspective of equality, diversity and inclusion in terms of risk exposure, perception and acceptance, and the linked notion of trust (see also Section 7.1.1), which ultimately might become the *lingua franca* for discussing the safety of complex systems rather than risk, at least in some sectors.

### 7.1.6 Theme six: Resilient complex systems

**Identify design-time and operation-time controls for increasing system resilience**

Design-time and operation-time controls can help to reduce safety-related risks. Current complex systems suffer from faults but normally they are successfully managed to ensure safety, as shown in the NATS example (see Appendix C.1.1). This example also demonstrated the role of preparedness and human factors in managing risk. This report views resilience as the system's ability to remain in a safe state despite unforeseeable events. Resilience is an operational concept but it needs design-time support to enable it, including for human oversight and control. Future work should develop ontologies of design- and operation-time controls for increasing resilience at the governance, management, and task and technical layers to provide practical guidance to designers and operators of future systems.

To achieve resilience, it is necessary to introduce controls across all the layers, including governance and management, not just task and technical issues. Future work should categorise existing methods to explain how they can be used in combination, using our framework as a structure (see the discussion of framework maturity in Section 3.5). Specific issues and questions to address include:

- **Governance layer:** There is a need to understand whether trying to regulate for resilience is more effective than regulating for safety or if it could be a useful complement. At a pragmatic level, this might mean reviewing and accepting safety management systems (SMS) to assess the extent to which they engender resilience principles. This will need to embrace sociotechnical issues, including human factors, at the management and task and technical layers.
- **Management layer:** Draw on good practice in contingency/emergency planning from mature sectors, for example clear command hierarchies – gold, silver, bronze; doing periodic exercises involving all the relevant stakeholders to build mutual understanding and experience of working together; and assessing the extent to which such practices need enhancement to deal with growing complexity. Again, the role of human factors is individual and group working is critical.
- **Task and technical layer:** Identify sources of gaps between intent and specifications in order to design the system and produce a safety assessment demonstrating how these gaps have been managed and minimised. This could include the use of modular approaches to design, where practicable, using interface control documents (ICDs) that enable safety properties to be managed effectively. Also, system-wide analyses should be performed where properties of interest cannot be localised, noting that the aim is to minimise the need for holistic analysis. These approaches should form part of an iterative development under change control, where designs are re-analysed based on the impact of the change in terms of uncertainty. It may also be possible to build on methods of agile software development, noting that there are several publications on agile safety-critical software on which to draw (with [122] being one of the most comprehensive).

A better integration of human factors engineering into the safety design of systems, especially for the those sectors focused on traditionally 'engineered' systems, will be required to ensure an adequate level of resilience not possible given state-of-the-art technologies and emerging system complexity. The role of the human operator in automated driving systems (see Appendix C.2.2) is a good illustration of these issues.

There is a substantial body of work on resilience including practical guidance for local governments such as [19], and more principled treatment of the concepts in the context of healthcare such as [73]. Resilience will remain of critical importance for future (even more) complex systems, but the challenges mean that more design-time and operation-time controls are needed to ensure and assure safety – probably including completely novel methods for *ad hoc* systems. It might be possible to further systematise resilience. Avenues to explore include developing guideword-based methods for identifying potential failures, as outlined in Section 3.5 and more explicit modelling of uncertainty, identifying *possibility* of events rather than probability, perhaps building on the Johari window, see Figure 16.

# 7.2 Research agenda

The study has considered a very broad range of research disciplines from safety engineering, complexity theory, social sciences and the law, as well as those that are domain-specific, of course. There will be a need to balance domain-specific and domain-independent research – and the **Safer Complex Systems** programme needs to consider this balance to seek to maximise the outputs. Further, there is likely to be benefit in the identification of grand challenges to focus research. We start by considering grand challenges before introducing key research areas and end with a discussion of methodology.

### 7.2.1 Grand challenges

In our view, the **Safer Complex Systems** agenda might be advanced by focusing on some grand challenges. To illustrate this, this report presents one domain-specific grand challenge (GC) (focused on mobility) and one that is sector-independent (focused on resilience, Theme six).

- **GC1: Last-mile delivery** – autonomous domestic delivery of food, medication and so on, often referred to as last-mile delivery, has potential societal benefits for health and wellbeing. These include supporting independent living for the elderly and infirm, and providing *resilience* in the face of future pandemics such as COVID-19. The primary challenge is to provide the capability safely, as the delivery system will need to operate on the roads, cross pavements and deal with a wide range of delivery situations, including multi-occupancy buildings. There are secondary challenges across all three layers: assurance and public trust at the task and technical layer, incident and accident analysis and supply chain issues at the management layer, and licensing and regulatory oversight at the governance layer.
- **GC2: The AI operator** – (human) operators play a vital role in ensuring resilience and safety of current complex systems. As complexity and autonomy grow it may be impossible for humans to play this role, so the primary challenge is automate resilience. This involves addressing the semantic gap in defining acceptable safe behaviour, operating within ethical and legal constraints so that the system meets societal expectations for safety, and developing ways of giving the system the common sense and problem-solving capabilities of an operator, so that it can respond to and manage risks, avoiding bias and achieving transparency in decision-making (respecting *equality, diversity and inclusion*), and gaining trust. There are secondary challenges across all three layers, including: assurance and contingency rehearsals at the task and technical layer, operational monitoring and change management at the management layer, and responsibility, liability and regulation at the governance layer.

These are intended to be illustrative, and it is expected that such grand challenges would be developed and refined with international stakeholders, see Section **7.3** below.

### 7.2.2 Key research areas

We believe that a holistic approach to managing safety of complex systems is a vital area of research that requires sustained and very large-scale support. Research bodies and funders should therefore scope and resource a major, multidisciplinary programme on **Safer Complex Systems** covering all the themes above. The following examples particularly focus on core safety concerns. This programme should address at least the following five key research areas (RA):

- **RA1: Design methods** (related to themes three, four, five, and six) – enhancement of systems and software design methods covering *inter alia*: modular design and use of approaches such as compositional design for assurance; inclusive design for complex systems, including ways of engaging a wide spectrum of stakeholders, such as use of VR, automatically generated explanations; making systems risk-aware to enable dynamic management of risk; complete lifecycle design including phasing out/replacing elements of the system, phasing out and migrating to a successor system; and global design, including ways of making designs sensitive to regional and cultural differences in risk perception, risk acceptance, and so on.
- **RA2: Safety analysis methods** (related to themes four, five and six) – new methods for analysing systems of systems at the governance, management, and task and technical layers, including: ways of incorporating systemic failures, for example those that arise from complexity not component failures; systematic use of generative methods, including those based on simulation, in support of safety assessment; agile methods for safety assessment of new combinations of systems (*ad hoc* systems of systems) allowing rapid deployment of novel configurations, building on and complementing ongoing work on safety assessment of AI and autonomous systems; and seeking to gain a good balance between the concepts and practices of Safety-I and Safety-II and drawing on methods such as FRAM.
- **RA3: Regulatory processes and legal framework** (related to themes one, two and three) – investigation of new regulatory mechanisms such as regulating for resilience and experimentation, for example via sandpits, to

assess the effectiveness of these new approaches; establishment of data exchange standards to enable effective incident and accident investigation; and identification of effectiveness of regulatory mechanisms, such as safety cases, for different types of systems and application to better inform the design of regulatory practices. This should also consider the limitations in current legal frameworks and identify whether significant changes are desirable,such as in the basis of tort law, or giving systems a limited form of personhood in a legal sense so that it is easier to apply the notion of vicarious liability.

- **RA4: Operational safety management** (related to themes three, four, five and six) – develop methods for improving resilience and safety, seeing them as complementary objectives – with resilience perhaps seen as more fundamental. This should include identification of leading indicators of problems to assist in managing risk; identification of how to use data analytics, such as ML, to analyse operational data to assess changes in risk and to predict failures; and investigation of the use of digital twins and simulation as the basis for operational controls and to support emergency planning and rehearsal for preparedness.
- **RA5: Resilience against malicious exploitation of system complexity** (related to themes four, five and six) – the topic of cybersecurity was identified as a threat to safety several times during this study. Examples include cyberattacks that use timing or power signatures of a chip to decode cryptographic keys, (also known as side-channel attacks [123]), and also manipulation of known insufficiencies in the systems (such as using manipulated traffic signs to trick ML-based

Engaged scholarship

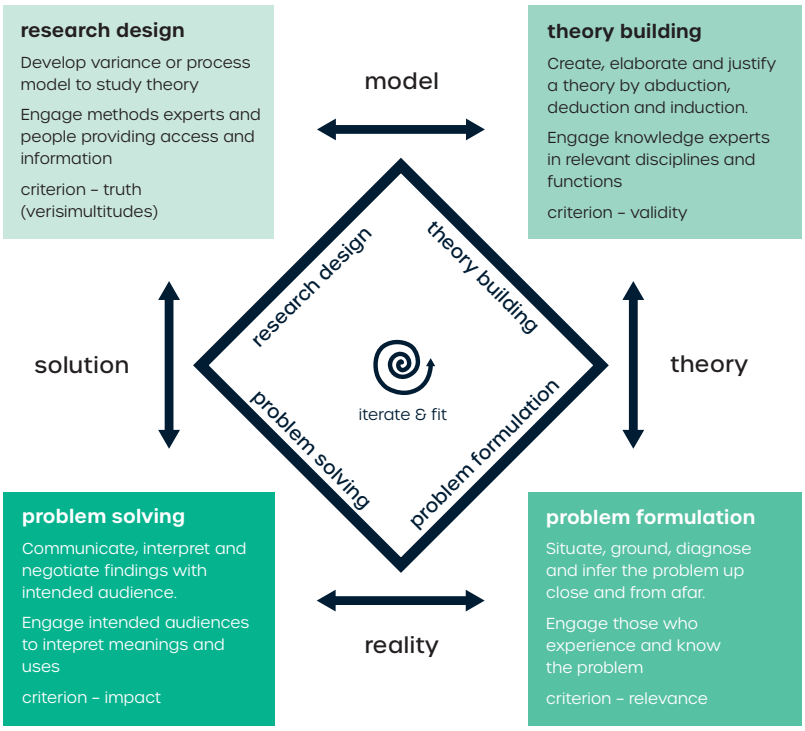



Figure 17: A model of 'Engaged Scholarship' from Van de Ven [126]

perception [124]). A common feature of these examples is that the attacker exploits emergent properties of the system for their own malicious gain. Research is therefore required to identify what causes of complexity lead to consequences that can be actively misused to cause a systemic failure and what design-time and operation-time controls can be applied to strengthen the systems' defences.

So far as practicable, these activities should be coordinated globally, recognising that different countries will have their own research priorities.

### 7.2.3 Methodology

Given the finding that existing safety management controls and governance capabilities are challenged by growing complexity (see Section 6.4) it is important to consider how we achieve confidence in new or refined safety methods.

Conducting research in safety is difficult and the rate of advance in methods has been surprisingly limited, given the advances in the systems themselves. There are many reasons for this (see, for example, [125]); the aim here is to shed some light on this from a methodological perspective, which should influence the way in which the remainder of the **Safer Complex Systems** programme is conducted.

One of the difficulties with safety engineering (and research in safety engineering) is that it is often effectively an open loop activity — assessing designs, but only being concerned about the effectiveness of methods when accidents occur [125]. This can be seen as a cultural problem in the safety community and one that is all the more exposed by the growth in system complexity. To address this, *evaluation* needs to be a core part of what the safety community does, especially in research, and

this would enable us to make more rapid progress and to be able to reject, rectify or replace unsound theories much more quickly than we do now [125].

There is perhaps a more profound shift needed — that of accepting that safety engineering is more like social science than science. The model of 'Engaged Scholarship' [126] (see Figure 17) seems to be much more representative of how work in safety engineering (research) is carried out than the scientific models of Kuhn [127] or Popper [128]. It also perhaps indicates why it is hard to transfer results between domains, because tangible progress is actually embedded in particular industrial or organisational settings so findings don't necessarily translate easily into other settings (let alone domains); the difficulty of adopting Safety-I practices in healthcare is a case in point. This also suggests that Lakatos' views of competing projects [129] is appropriate for safety as evidence from one domain (even counter evidence) is not necessarily very compelling in another, so it is very hard to accumulate compelling evidence to replace a theory. The notion of a self-critical and learning community, which evaluates what it does using the sorts of engagement schemata outlined in Figure 17, is probably a better model for making progress in safety engineering than those used hitherto – and perhaps should be the basis for a Safety-III.

For the **Safer Complex Systems** programme to be effective it is important that it adopts an appropriate (research) methodology. It is suggested that serious consideration be given to adopting a model such as Van der Ven's 'Engaged Scholarship' and a focus on evaluation as a core principle for implementing the programme's research agenda, see recommendation SCS4 below.

## 7.3 Future directions for the Safer Complex Systems programme

This report has indicated a direction for work on **Safer Complex Systems** based on a holistic approach to considering factors impacting risk across the governance, management, and task and technical layers. Based on the findings of the study, we defined several long-term themes for future work in Section 7.1, and set out a research agenda, with a focus on safety, in Section 7.2. This report concludes with more specific guidance for the next steps in the programme itself.

We recommend that Engineering X addresses the following during the remainder of the **Safer Complex Systems** (SCS) programme.

- **SCS1: Framework validation and refinement** – work with regulators, industry bodies and others to identify 'case studies' in a range of domains, including built infrastructure, and large white goods (see Appendix C.10), that might benefit from use of the framework presented in this report as a *descriptive* tool to help both in validation and to provide insights that enable the framework to be refined.
- **SCS2: Framework enrichment** – the framework should be enriched so it can be used for (safety) analysis by integrating new or existing models underpinning each of its areas, and enabling it to deal with both designed systems and supporting rapid assessment of *ad hoc* systems; This should embrace *inter alia* systems engineering methods, risk communication and resilience engineering, and should reflect the ideas of developing and maturing the framework set out in Section 3.5.
- **SCS3: Global regulatory collaboration** – establish a global consortium of regulatory and similar bodies, such as ICAO, IMO, UNECE and WHO, to focus on domain-specific issues. This should include reviewing and refining the domain-specific recommendations (see Appendix D) to identify domain-specific grand challenges, to seek to identify funding sources for these challenges and to help coordinate across domains to maximise the benefit that can be gained from shared understanding, regulatory approaches, and so on.
- **SCS4: Global research collaboration** – establish a global consortium of research funders including charities, such as Lloyd's Register Foundation, and government bodies, such as the US National Science Foundation (NSF) and UK Research and Innovation (UKRI), to ensure coverage of the research themes and agenda (see Sections 7.1 and 7.2) and to encourage international research collaboration. This should identify, promote and support some sector-independent grand challenges perhaps using some of these as a basis for establishing an ARPA-like initiative in the UK. A key success factor in implementing the research agenda will be to adopt an appropriate methodology (see Section 7.2.3).
- **SCS5: Competencies** – it is very apparent that managing the safety of complex systems requires skills across a range of disciplines, including complexity science, safety, systems engineering, organisational design, human factors engineering and the law. Thus carrying out such work requires multidisciplinary collaboration, but also broadening the skills of researchers and practitioners to improve their ability to work together. Work is needed to define, then implement, education and training programmes (including PhD level) that provide a broad understanding of **Safer Complex Systems** for all participants but going into depth in at least one, and preferably two, specialist disciplines as appropriate for each participant. This should also address specific issues such as the skills to select appropriate sets of safety analysis methods (see Section 4.5); to do this is likely to require collaboration across educational institutions, such as is done in European Training Networks (ETNs).
- **SCS6: Review of scope and strategy** – the study deliberately took a view of 'safety' that considered harm to humans, although it did consider how harm could arise indirectly, via environmental effects such as pollution for example. The Technical Advisory Group (TAG) questioned whether the scope should be broadened to include environmental, economic/business and long-term effects such as global warning. The TAG also suggested that there might be merit in embracing a 'paradigm shift' to acknowledge the impact of system complexity on safety in all domains – and in bridging domains that have previously been considered independent, as highlighted by COVID-19. This would include, for example, considering the differences between short-term or immediate risk related to a system, and the longer-term consequences of deploying such systems. Against this background, a decision should be made whether the benefits of expanding the scope of work as outlined above outweigh the difficulties of making useful and practical progress – so that any scope changes are made with a proper understanding of their consequences.

The scope of the **Safer Complex Systems** programme is already very broad and its objectives will not be achieved by the Academy alone. Should the scope be expanded, then it is all the more apparent that the Academy alone cannot address

all the challenges. Thus, regardless of the scope and detailed research and policy agenda that is adopted, international cooperation will be a key to enabling the programme to deliver.

This report indicated a direction for work for Safer Complex Systems based on the framework presented. Suggestions include validating, refining and enriching the framework, collaborating across the globe on grand challenges, working to broaden the skills of those working on complex systems, and reviewing the scope and strategy of the Safer Complex Systems programme.

# 8

# Conclusions

The increasing complexity of the systems upon which our society depends is challenging our capabilities in safety engineering. We must evolve rather than fully transform our approach to the assessment and management of safety. This section gives an overview of the work required at all three layers in the framework to help us to do this.

# 8.1 Conclusions

The challenge of complex systems is enormous. This report has been produced at a time when COVID-19 has affected most nations on the planet and has shown clearly the dependencies between systems that were conceived as being independent. Complexity is challenging – outstripping – our capabilities in safety engineering and we may be at a 'tipping point' where the levels of safety that society has come to expect may not be sustainable. More starkly, the safety community does not have good methods for assessing and managing safety of *ad hoc* systems. Further, a change in focus is needed to encompass the indirect effects of systems through unplanned system interdependencies and through environmental effects, such as air pollution, on human health and safety.

Some argue that a paradigm shift is needed in our approach to assessment and management of safety. Perhaps this is partly true – but to dispense with the accumulated experience of decades of successful safety engineering of systems as complex as aircraft, cars and nuclear power plant, is unlikely to be beneficial. Instead what is needed is *both* an improvement in established techniques to help them to scale and to address engineering causes of complexity, and a more radical and innovative approach to dealing with *ad hoc* systems and the unplanned interdependencies between them.

To do this successfully requires work at all three layers in the framework. In governance, there is likely to be a need for changes in regulatory structures (such as reducing the number of regulators), developing new approaches to regulation, perhaps focusing on controlling safety management competency in organisations, as opposed to focusing on the systems themselves. There may also be a need for changes in legal structures, perhaps introducing limited notions of personhood for autonomous systems (making them 'legal individuals' in a narrow sense) and changing the basis of tort law.

At the management layer, there will need to be more of a focus on operation-time controls, and designing systems to enable operational control – even where the role of individual systems in a wider system-of-systems can't be anticipated at design time. The role of humans in managing safety will be crucial as they are often the key factor in achieving operational resilience – but there will also be merit in considering how automation can help, as systems become so complex that individuals cannot achieve sufficient situational awareness.

At the task and technical layer, the advances needed are both to do with the advances in technology – AI/ML, autonomy, IoT, to name but a few – and the intricacies of human-system interaction. Perhaps this is where the focus should be on extending existing methods of analysis – to complement more radical changes at the governance and management layers. However, consideration also needs to be given to design approaches as some traditional safety architectures, such as using redundancy and diversity (in the technical sense), are challenged by the emerging technologies. Overall a balance is needed between improving 'what worked' in the past and radical innovation to address the challenges of complexity.

Further, the programme needs to embrace equality, diversity and inclusion – both to ensure equity in terms of risk distribution and to embrace the benefits that accrue from diversity of thinking in managing and governing systems. This should be a core theme of future work but the focus should be on its intersection with the engineering of complex systems – considering, for example, societal causes of discrimination should be seen as outside the scope of the programme.

There is attraction in conducting research, and other activities in the programme, in a domain-independent way – as results gained can, in principle, be applied widely across domains. However, there are many domain-specific constraints, for example the risk focus on individuals in healthcare as opposed to a greater focus on populations in other domains, the need to get community acceptance of risk targets, and so on, which means that domain-independent work has its limits. It is suggested that the primary focus for the rest of the programme should be on domain-specific activities, complemented by domain-independent work where appropriate; guidance on human factors engineering might be a good candidate for domain-independent work.

During the study it has become very clear that complex systems can have many, and far reaching, consequences – on business, the economy, the environment, and so on. This has been highlighted by COVID-19 but is also apparent in smaller scale incidents, such as extended loss of electrical power to a community, and in telecommunications outages. There is a temptation to expand the scope of the study to these broader issues – but we would like to sound a note of caution. If the same (or strongly related) analysis methods can shed light on all the consequences of interest, then there is great merit in adopting such a broad view. However, the nature of dependencies in economic and business systems, for example, are quite different from those in technical systems, and the nature and acceptability of risk is also very different. There is, of course, merit in seeking to understand if similar

analysis methods can be applied across a range of concerns – but the challenges and *importance* of safety are sufficiently great that we recommend a continued focus on safety itself, to avoid the risk of dissipating effort by addressing too broad a problem.

Finally, it is important to consider the methodology for conducting the rest of the **Safer Complex Systems** programme – including the research elements. As complex systems are long-lived and (generally) have low accident rates it is hard to evaluate the effectiveness of safety assessment and management methods directly (although there are some surrogate methods). Thus, an alternative approach is needed that enables evaluation and learning to be built into the programme in a more integrated way – the 'Engaged Scholarship' model from social/management science [126] may be one appropriate methodology. It certainly is well-attuned to the conduct of domain-specific activities and thus could have an important role to play in the future of the programme.

Complexity is challenging our capabilities in safety engineering. We need both an improvement in established safety engineering techniques to help them to scale and to address engineering causes of complexity, and a more radical and innovative approach to dealing with *ad hoc* systems and the unplanned interdependencies between them.

# A

# Definitions of terms used in the study

This section gives a definition of safety and complexity, as well as defining the terminology used in the framework.

# A.1 Definition of a system

The following terms are used to describe structural characteristics of systems:

- **System** – An arrangement of parts (or elements) that together exhibit behaviour or meaning that the individual constituents do not.
- **Element or part** – A component or constituent of a system, noting that this may be a system in its own right (see system of systems).
- **System of interest** – The scope of the system under consideration is defined in terms of the boundaries between the system under consideration and the environment as well as the set of objectives that the system shall fulfil.
- **Interconnectivity** – Elements or parts of a system that can interact or communicate, for example by sending signals on a digital network or through physical linkage.
- **Interdependency** – Behaviour of one element depends on that of another, although there is no connectivity.
- **Open system** – A system that has flows of information, energy and/or material between the system and its environment, in and out of the system boundary. The system can adapt to the exchange. System boundaries are therefore often referred to as fuzzy, or semi-permeable, and can be difficult to precisely define.
- **Systems of systems** – A collection of component systems that can exist independently and have different owners who may have conflicting objectives.

For example, a car is an arrangement of tyres, brakes, wheels and so on that exhibits behaviour including mobility, load-carrying, that the parts do not, hence is a system. The parts of cars are interconnected. Road traffic is a system including roads, lorries and cars, which exhibits behaviour that the parts do not, for example queues and 'shockwaves' in motorway traffic. The parts (or elements) of road traffic are interdependent, but generally not interconnected (although that may change as technology evolves); thus, road traffic can be seen as a system and also as a system of systems.

Systems with a particular set of complicating factors are often referred to as 'Wicked Problems'. This describes situations where our ability to intervene in complex systems is sometimes limited by both the context in which the system exists, and the features of the system itself [6, 130]. 'Wicked Problems' are also the result of complex systems with poorly understood, defined, or contradictory requirements.

# A.2 Definition of complexity

It is common to distinguish between complicated and complex systems. General systems theory talks about the possibility to predict all possible behaviours of the system even when the number of components and interactions is large [131]. In the terms of complex systems theory this is complicated, not complex, as there is no emergent behaviour. From a different perspective, a complicated system, for example an aircraft, will not change its intended behaviour by itself (it will always be an aircraft although others might change it from carrying passengers to carrying cargo). However, a complex system, such as a biological organism or a multinational company, might change its intended behaviour; in the case of a company this might be from selling products to selling services.

From the perspective of complexity science there are several characteristics that are shared by most, if not all, complex systems. These are variously described and defined in [2, 4, 5]. These are introduced here with some small terminology changes for ease of understanding by an engineering audience:

- **Complexity** – A system is complex if some of its behaviours are emergent properties of the interactions between the parts of the system, where you would not be able to predict those behaviours from knowledge of the parts and their interactions alone.
- **Unity** – Aristotle describes things with unity being those "which have several parts and in which the totality is not, as it were, a mere heap, but the whole is something besides the parts", Aristotle's Metaphysics [132, pages 8–10].
- **Emergence** – The observed behaviour of the system cannot be predicted from knowledge of the parts and their relationships (this reflects Aristotle's definition of a system).
- **Self-organisation** – The internal parts, through interactions between them and the environment, will arrange themselves to produce emergent global system behaviour with no central control system.
- **Autopoiesis** – The system state is resilient to external shocks and loss of internal system components. This is also linked to the concept of system inertia, where the impact of shocks or interventions may not become observable immediately.
- **Non-linearity** – Small perturbations may have a proportionate response, no response, or a disproportionate response (which might include a change in intended behaviour).
- **Coupled feedback** – The system continuously cycles through similar states and there is both internal feedback and feedback via the environment where the output of the system forms, or influences, part of the input and the feedback paths can influence each other.
- **Mode transitions** – Complex systems can go through changes in mode, where the systemic behaviour of a system radically changes without an easily discernible trigger or change in the environment (an example is the onset of 'shockwaves' in motorway traffic).
- **Boundaries** – Complex systems are said to be open, and things flow in and out of them; they therefore have semi-permeable boundaries that are often fuzzy and difficult to define. Although the treatment of 'boundaries' is standard there has to be a flow in and out of a system, for example of light or electricity, for it to be able to interact with its environment. A possible refinement is to say that matter can flow in and out of open systems but not closed systems, hence the constituent parts of open systems can evolve over time.
- **Inertia** – Shocks to, or interventions in, a system may not produce an immediately observable effect, because of a time lag in which no effect can be observed. This can make it difficult to know if an observable change in the system state is in response to an intentional intervention, such as a policy change, or something else that happened at another time.

## Characteristics of complex systems

### Unity

A set of interconnected parts, where the whole is greater than the sum of the parts. This includes systems that are explicitly engineered or that form in an *ad hoc* manner.

### Emergence

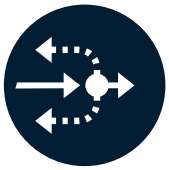
Observed behaviour of the system cannot be predicted from knowledge of the parts and their relationships.

### Semi-permeable boundaries

Things (information, energy, matter, people, etc.) can flow in and out of the system, boundaries change and are hard to define.

### Non-linearity

Similar perturbations or system inputs may have a proportionate response, no response or a disproportionate response.



### Coupled feedback

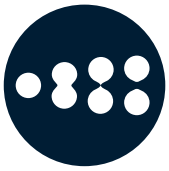
There is both internal feedback and feedback via the environment where the feedback paths can influence each other.

### Mode transition

Systemic behaviour of a system radically changes without an easily discernible trigger or change in the environment

### Self-organisation

The internal parts, through interactions between them and the environment, will arrange themselves to produce emergent global system behaviour with no central control system.



### Autopoiesis

The system is resilient to external shocks and loss of internal system components with the ability to reproduce and maintain itself

### Inertia

Shocks to or interventions in a system may not produce an immediately observable effect, due to a time lag in which no effect can be observed

Figure 1

# A.3 Framework terminology

The following is an attempt to organise the perceived characteristics identified by the study team, and those arising from the stakeholder workshop on 14 February 2020, into causes, consequences and systemic failures. It also includes the other three areas of the framework, namely exacerbating factors and the design-time and operation-time controls for managing complexity safely. Where practicable, the terms are illustrated with examples from the report or from relevant literature.

### A.3.1 Causes

Characteristics that can be viewed as causes of the defining characteristics at the governance layer:

- **Multiple jurisdictions** – Parts of the system are in different (potentially incompatible) legal or regulatory jurisdictions, such as countries, or at sea and on land, see the brief discussion of maritime regulation in Section 3.4.
- **Diversity and inclusion of needs and risk perception** – Different actual risk distribution and perceptions of risk between different groups, for example the impact of living conditions on spread of disease such as COVID-19 (see Section 5.3).
- **Rapid technological change** – Technology is adapting much faster than safety standards and regulations potentially leading to unnecessarily complex solutions to comply with standards and incorporate the new technology, or increased risk due to inadequate control of the new technology, such as use of AI or ML (see Section 3.4).
- **Weak science basis** – Inadequate knowledge on how to predict system behaviour and/or to assure safety of some technology or system design, for example the limited understanding of the transmission of COVID-19, why some people are asymptomatic, which makes predictive modelling difficult.

Characteristics that can be viewed as causes of the defining characteristics of the management layer:

- **No single owner** – Different legal individuals own and/or are responsible for different parts of the system, for example thousands of organisations were involved in the Haiti disaster relief, and this posed problems for coordination and even evaluation [117].
- **Supply chains and cross-domain collaboration** – Supply chains can be very complex (really networks not chains); they change frequently and may involve suppliers from other domains working to different standards giving rise to problems of materials integrity, such as the E. Coli contamination discussed in Appendix C.4.1.
- **Diversity and inclusion of stakeholders in design and operation** – Diversity, particularly, can improve the quality of decision-making, and being inclusive ensures that all relevant viewpoints are considered, for example there is value in 'cognitive diversity' in achieving collective intelligence in decision-making [133].
- **Path dependency** – Systems are a product of, and are constrained by, their history and how they have evolved, for example the NATS system that failed in December 2014 included legacy elements (see Appendix C.1.1).

Characteristics that can be viewed as causes of the defining characteristics of the task and technical layer:

- **Heterogeneity of system components** – The more types of system element there are, the harder it is to analyse the system as they have different properties, such as in healthcare, involving patients with different conditions and comorbidities, clinicians with different skills and experience and many different technical components: see, for example, the sepsis accident in Appendix C.3.1 (note that some classes of system can be treated statistically, but many complex systems are sufficiently heterogeneous to defeat simple analysis, but not large enough that statistical techniques are effective).
- **Interconnectivity and interdependency** – Connectivity/dependencies do not form a simple regular structure, are typically many-to-many, and may involve cyclic dependencies, such as the interdependencies of road traffic and roadside infrastructure, see the discussion of system in Appendix A.1.
- **Memory** – Prior states influence the current and future states, for example the NATS system stored and 'replayed' inputs causing a redundant system to fail in exactly the same way as the first (see Appendix C.1.1).
- **Environmental complexity and open system boundaries** – The environment is hard to bound, dynamic and may change fundamentally over time, in the form of new element types such as autonomous air taxis.
- **Human-system interaction** – Operators or users interact with the system to make or confirm key decisions, for example some of the Watchkeeper accidents involved a mismatch between the operators' understanding of the system state and the true system state [134].
- **System evolution, adaptation and self-organisation** – Composition of the system, properties of individual parts and the interconnections or interdependencies change over

time (without central control), such as the composition of traffic on the road.

- **Collaborating non-hierarchically managed systems** – Parts of the system work together pairwise or in groups to meet (shared) objectives, for example cars/drivers in dense urban traffic share road space so as to make progress and avoid collisions.

Note that memory and path dependency are different – path dependency is about how the system design and composition has arisen, whereas memory is about system behaviour. Both apply in the case of the NATS failure, but they are different contributory causes to the ATM outage.

### A.3.2 Consequences

Characteristics that can be viewed as consequences of the defining characteristics of complex systems at the governance layer:

- **Competing objectives** – Regulators or other stakeholders of a domain have conflicting objectives and potentially inconsistent standards or regulations, for example the state of Arizona wished to promote autonomous driving that, with hindsight, can be seen to have been in conflict with safety objectives (see Appendix C.2.2).
- **Competency gaps, standards and regulation lag** – Complex systems, and particularly new technologies such as AI and ML, are unfamiliar to many regulators; standards and regulations often take a long time to produce, meaning that they 'lag behind' the state-of-the-art, for example there are no regulations for operations in the stratosphere (above 60,000 feet, or FL600), which several companies are seeking to exploit to provide telecommunications services to remote areas, see for example [135].
- **Accountability and moral responsibility gaps** – Inability to 'hold to account' any regulator, as overall responsibility for safety is unclear or the conditions for moral responsibility are not met, for example new technology enables new classes of products that are not covered by any existing regulator or regulation, such as seems to be the case with unmanned stratospheric craft.

Characteristics that can be viewed as consequences of the defining characteristics of complex systems at the management layer:

- **Accountability mismatch** – The legal framework is such that none of the stakeholders who are responsible for the system design and operation can be held to account for failures, for example developers of a system not taking a 'system authority' role with overall responsibility for safety and relying on contractors to collaborate and to coordinate to ensure the safety of the overall system.
- **Competency gaps, unmanageable complexity** – The system has a level of complexity that makes it impossible for those responsible to manage (with confidence) to ensure safety, such as COVID-19 (see Appendix C.3.2).
- **Risk transference** – Risks are transferred between stakeholders without them and other stakeholders necessarily being aware that this is happening, for example the Uber Tempe accident risks were transferred to the pedestrian and the safety driver (see Appendix C.2.2).
- **Accountability and moral responsibility gaps** – Inability to 'hold to account' the designer or operator, as overall responsibility for safety is unclear or the conditions for moral responsibility are not met, for example the state of Arizona stating that Uber had a case to answer following the fatality in Tempe (see Appendix C.2.2).

Characteristics that can be viewed as consequences of the defining characteristics of complex systems at the task and technical layer:

- **Coupled feedback and inertia** – The system self-stabilises so that external 'shocks' are absorbed and the system behaviour is minimally disrupted,

For example an aircraft engine when flying into adverse weather conditions such as a rainstorm continues to deliver the demanded power and thrust (this is a characteristic of control systems in general).

- **Non-linear behaviour** – Small perturbations may have a proportionate response, no response, or a disproportionate response, such as the transition of a crowd of people from a peaceful demonstration into a riot.
- **Semantic gap** – The 'gap' between the intent of the system and its specification, which makes design, verification and validation difficult, for example for an autonomous vehicle there is a significant difference between the 'intent' of avoiding collision with pedestrians and the specified (specifiable) behaviour in terms of recognising pedestrians from images, predicting their trajectory and taking avoiding manoeuvres, if necessary.
- **Non-determinism and emergent properties** – The observed behaviour of the system cannot be predicted from knowledge of the parts and their relationships,such as the load-carrying capability of cars, see the discussion in Section 2.
- **Mode transitions (or tipping points)** – A sudden and radical change in system behaviour arises without obvious major change in inputs, such as the onset of 'shockwaves'

in motorway traffic, see the discussion of system in Section 2.

Arguably, emergence underpins all these perceived consequences. This is unsurprising as emergence can be viewed as the most fundamental of the defining characteristics of complexity.

### A.3.3 Systemic failures

The identification of systemic failures is one of the more novel aspects of the framework, and it is the area that is expected to be subject to most revision in later phases of the **Safer Complex Systems** programme.

Systemic failures that can be viewed as arising from the characteristics of complex systems at the governance layer:

- **Inappropriate deployment decisions** – Beneficial systems cannot be deployed because society or key stakeholders are not ready for them (or *vice versa*, such as systems deployed that are not beneficial), for example difficulties in getting acceptance for novel technologies in healthcare, such as AI, where clinicians retain responsibility for the outcome of using the technology with limited visibility and control.
- **Inadequate regulatory control** – Regulators do not impose sufficient control over the system allowing an unsafe design to be deployed or to continue to operate, for example the approval of the 737 MAX by the FAA with limited oversight and failure to recognise weaknesses in the safety process [136]; or the safety case for the Nimrod XV230 where the aim was to provide a safety case for a system assumed to be safe (due to its long period of operation) rather than to expose the true risks [137].

To illustrate the links between the consequences in the framework and systemic failures, the two examples above of inadequate regulatory control can both be seen to arise from competing objectives, where commercial objectives over-rode safety concerns. In the Nimrod case, this traces back to path dependency, among the causes in the framework, as Nimrod was a legacy aircraft whose design was not fully understood at the time the safety case was produced.

Systemic failures that can be viewed as arising from the characteristics of complex systems at the management layer:

- **Accountability mismatch** – The legal framework is such that none of the stakeholders who are responsible for the system design and operation can be held to account for failures, such as the Uber Tempe fatality where Uber were found to have no case to answer under Arizona law (see Appendix C.2.2).
- **Inequitable risk distribution** – The design or operation of the system is such that (safety) risks are shared inequitably between different groups (this is an aspect of diversity and inclusion), for example the impact of COVID-19 on ethnic minorities and poorer populations (see for example Section 7.1.3).
- **Unanticipated risks** – Systems pose risks, or classes of risk, that were not anticipated during the development of the system, for example the impact of the addition of external cladding to Grenfell Tower, contributing to a catastrophic fire that was not anticipated at the time the cladding was added [138].

Unanticipated risks might arise because of disparity between the maturity of a technological application and the preparedness of development and operational organisations to adopt, and adapt to the changes brought about by that innovation – in other words 'Competency gaps' at the management or governance layer.

Systemic failures which can be viewed as arising from the characteristics of complex systems at the task and technical layer:

- **Model mismatch** – The world model held by the system is different to that held by operators, other collaborating systems in a system of systems, and/or that used at design time, for example the A320 accident at Okęcie, Warsaw, where the pilots 'landed' the aircraft in bad weather, but the aircraft logic determined that the aircraft was still flying and did not deploy the ground-braking systems for about 10 seconds [139].
- **Authority mismatch** – People who have responsibility for actions (to ensure safety) do not have the authority to discharge those actions because of the design of the system, for example the pilots in the 737 MAX accidents had limited authority against MCAS and the aircraft aerodynamics (see Appendix C.1.2).
- **Decision mismatch** – Decisions or recommendations are made on available data that are not appropriate (safe) given a fuller understanding of the context, such as the decisions made in the treatment of sepsis (see Appendix C.3.1).

Some level of 'mismatch' occurs with all systems, as models held by the system are always partial views of the environment and lag behind the environment because of sensing and processing delays (human or automated), so the above should be viewed as meaning mismatches to an extent that is 'safety-significant'.

### A.3.4 Design-time management controls

Design-time controls for managing complexity safely at the governance layer include:

- **Normative/outcome-based standards** – Where technology (or other relevant factors) change fast, then outcome-based, or goal setting, standards can be effective. They go out of date less quickly since they define targets to achieve not specific tasks to be undertaken, although they require more skill to interpret than prescriptive (rule-based) standards.
- **Legislation** – Formal legislation can both provide incentives for compliance and guard against (prohibit) system designs that would fall into the 'Unknown' region of the Johari window.
- **Tort/common law and soft law** – Tort or common law provides a basis for prosecutions where rights have been infringed, and this may apply regardless of system or technology. Further, so-called 'soft law' – rules adopted voluntarily by an industry or sector – may enable fast, industry-wide treatment of issues not addressed through formal regulation.
- **Diversity and inclusion in policy and regulation** – There is evidence that diversity improves decision-making [133] and diversity in formulating policy and regulations should help avoid unconscious bias and produce outcomes that are fairer and do not result in unequal risk distributions.
- **Engagement in development** – It is difficult to assess complex systems as a 'product' and regulatory engagement in development is a way of gaining system understanding, which would not be available to an end-of-development assessment; note that this is standard practice in aerospace but, as the 737 MAX accidents show, such practices can also 'fail' (see Appendix C.1.2), and are not practical for *ad hoc* systems.
- **Publicly available specifications** – Standards bodies are developing Publicly Available Specifications (PAS), typically over about one year, enabling a rapid response to new issues en route to standardisation, for example the British Standards Institution (BSI) is producing PAS for autonomous vehicles [17].
- **Community guidelines** – Professional communities can develop industry guidelines for dealing with emerging technology, for example the Global Mining Guidelines Group has developed guidance for autonomous systems in mining and quarrying [18], and this enables the industry to move rapidly on a consensual basis where formal regulation moves slowly.
- **Learning from experience** – While the (systemic) failures of complex systems can be unprecedented, often there are similarities with previous events (compare SARS and COVID-19) and individual causal factors will often have been seen previously, so learning from experience allows steps to be taken to avoid recurrence of similar events – good practice would suggest learning before, during and after events [117].

Design-time controls for managing complexity safely at the management layer include:

- **Stakeholder engagement** – Involvement of all classes of stakeholder in development, particularly in requirements elicitation and in establishing acceptability of risk, increases the chance of achieving system safety and that it will be acceptably safe for all.
- **Safety management system** – A safety management system (SMS) is a systematic approach to managing safety, including the necessary organisational structures, accountabilities, policies, and procedures. An SMS provides a systematic way to identify and control risks, as well as providing assurance that risk controls remain effective and legislative requirements are met. Effective SMS implementation includes an element of continuous improvement where the organisation monitors and assesses the effectiveness of their SMS to enable enhancement of safety management practices.
- **Voluntary codes of practice** – Voluntary codes of practice influence organisations and set benchmarks for acceptable practices. They embody agreed good practice and provide a means of self-regulation for organisations.
- **Principles of high reliability organisations** – The following principles are ideally embedded in high reliability organisations (HRO): preoccupation with failure, reluctance to simplify, sensitivity to operations, commitment to resiliency, and deference to expertise. A further discussion of HROs is presented in Section 6.4.5.
- **Active risk management** – Risk management is not a static exercise conducted once. Risk assessments should be updated as new information becomes available and in response to change in the system design or the operating environment. Conduct of risk assessments should be a participative activity involving key stakeholders actively engaging with the process.
- **Change management** – Safety change management defines a process to identify changes that may affect the level of safety risk of a system and to identify and manage the safety risks that may arise from those changes. It should be integrated with other change management activities to ensure it is conducted at the most effective point in the process.
- **Agile development** – An

approach to producing systems and software so as to be responsive and adaptive to a rapidly changing environment. There are many different approaches but they are usually based on 12 principles from the Agile Manifesto such as "welcoming changing requirements even late in the development process" [140] with some also embracing safety within the agile framework [122].

- **Incremental delivery** – Provide system and software capability progressively (in small increments) allowing for early stakeholder validation and thus helping to reduce the 'semantic gap' [16]; this is strongly related to the ideas of agile development, and the principle of 'delivering working software frequently'.
- **Decision rationale** – Providing reasons for key design decisions so that their significance can be properly understood when responding to requirements for change (noting that many faults and failures arise from incompletely understood change) [141].
- **Diversity and inclusion in decision-making** – See the governance layer definition.
- **Supply chain/network management** – Overseeing the supply network to ensure that safety-related elements are identified, traced and managed, for example provenance of critical components, bi-directional flow of safety-relevant information and, where appropriate, risk-informed decision-making to maximise the ability of the supply chain to deliver a safe system as a collective.
- **Competency management** – A 'formal' process for defining and achieving the mix of competencies needed in an organisation to ensure that the (safety) skills required are developed and maintained, for example [142] (although this code of practice would need to be updated to deal with safety of complex systems).
- **Learning from experience** – See the governance layer definition.
- **Safety/assurance cases** – The decision to enter a system into operation requires management (and regulators) to be assured that the system is acceptably safe to operate. A safety or assurance case is a structured argument and evidence that can provide sufficient assurance to decision-makers that this objective can be achieved.
- **CONOPS** – The Concept of Operations (CONOPS) describes the (proposed) system in terms of the user needs it will fulfil, its relationship to existing systems or procedures, and the ways it will be used to foster consensus among stakeholders.
- **System integrator** – An authority tasked with drawing together the different parts of a system and making them work together effectively, including overseeing all aspects of safety so that individual parts are safe in themselves and the whole system meets safety requirements.

Design-time controls for managing complexity safely at the task and technical Layer include:

- **Diversity and redundancy** – Redundant architectures, such as command-monitor, duplex and triplex, are well-established and still applicable to elements of complex systems. Diversity, in the technical sense, of providing system parts with different functional or physical implementations to meet the same goals, can help to address sources of emergent properties and thus systemic failures; however, achieving diversity in systems using machine learning currently seems to be beyond the state-of-art.
- **Risk and hazard analysis** – Apply 'classical' risk and hazard analysis methods to systems (see Section 4) early and continuously through their life, particularly assessing proposed changes before they are implemented, recognising that these will need to be enhanced to deal with the failure modes of complex systems (see Section 3.5).
- **Design for assurance** – Design systems so that their functions and, in particular, structure make them easy to analyse and assure. For example, using modularity, contract-based design where the interface contracts include safety properties such as failure-signalling and programmed response to failures. To be effective the focus on assurance has to be a critical factor used in design reviews.
- **Independent assessment** – Employ a third-party, independent from the main design time, to assess the design (and other development artefacts) from a safety perspective, to provide independent assurance of safety and/or to expose weaknesses in the design that have been overlooked by the developers. This is established good practice in some domains, such as the rail sector.
- **Design for cyber resilience** – Design systems so that they are resilient to cyber attacks and can provide continued service, recognising that the connectedness of Cyber-Physical Systems (CPS)/the Internet of Things (IoT) provides a significant attack surface [143] [144] and noting that there is also a need to consider the impact of security on safety [99].
- **Inclusive design** – To make the system usable by, and safe for, all, systems must be designed for all users, not just the 'average'. Systems should be designed for extremes, such as the elderly or

infirm, ensuring that the diversity of users is considered.

- **Standards compliance** – Complying with applicable standards can assist in ensuring safety of systems and system parts. For example, for guarding hazardous machinery there are now conflicts between standards and complex systems, such as requiring physical separation, which can limit the utility of collaborative robotics (cobots). There is a need for intelligent interpretation and application of standards, see, for example [145], for a discussion of the relationship between safety assurance and standards in manufacturing in the context of the fourth industrial revolution.
- **Simulation and modelling** – Simulation, sometimes known as synthetic environments, and other forms of modelling are useful in obtaining understanding of system and environmental properties early in the development process and hence guiding the design to ensure safety. Simulations can also be 'driven' to explore safety properties to ensure designs are robust, for example assessing air vehicle sense-and-avoid algorithms [146]. The use of digital twins can enable such capabilities to be extended through life, see operation-time controls.

### A.3.5 Operation-time management controls

Operation-time controls for managing complexity safely at the governance layer include:

**Incident and accident analysis** – To prevent future safety occurrences, investigation and analysis of incidents and accidents is a key component of good safety governance activities. A structured approach for individual event investigations or more systemic analysis activities will ensure a more robust result to facilitate the learning of lessons. Trend analysis of events across multiple organisations or within a single organisation may provide additional insight into safety performance, which may indicate the possibility of future safety issues.

- **Legislation** – See the design-time definition.
- **Tort/common law and soft law** – See the design-time definition.
- **Diversity and inclusion in public engagement** – The role of the governance layer in representing societal values means that it is necessary to ensure diversity in public engagement and inclusion of as broad a range of stakeholder viewpoints as is practical when conducting governance layer activities. This is particularly important in legislation and guidance development and must be undertaken in a proactive manner to ensure all societal contexts are considered.
- **Operational monitoring** – Regulators should conduct operational monitoring of organisations to ensure compliance with legislation and that compliance is delivering the intended safety results. Operational monitoring at the governance layers is key to knowing whether intended legislative safety outcomes are being achieved by industry.
- **Active alerting** – Mechanisms to provide immediate notification of unsafe systems or services to the regulator or safety incidents to relevant governance organisations (such as search and rescue or investigatory authorities) allows timely intervention by relevant authorities in the management of safety occurrences or potential future accidents. Similarly, mechanisms to provide timely safety-relevant information from regulators to organisations (for example, faulty equipment alerts) provides mechanisms for safety knowledge to be shared among stakeholders quickly and effectively.

Operation-time controls for managing complexity safely at the management layer include:

- **Incident and accident analysis** – To prevent future safety occurrences, investigation and analysis of incidents and accidents is a key component of good operational safety management practices. Organisations should conduct their own investigation and analysis activities at a more granular level than governance organisations. Organisations should expect to investigate and analyse safety occurrences that governance organisations do not have capacity to review. Events analysed do not have to involve actual negative safety outcomes (such as injury or loss of life) - as much can be learned from minor safety occurrences or near misses.
- **Safety management system** – See the design-time definition. Safety management activities continue through the life of the system to ensure acceptable risk levels are maintained during operation and decommissioning.
- **Monitoring and analysis** – Safety performance monitoring allows an organisation to verify the safety performance of a system and validate the effectiveness of risk controls. Through-life monitoring and safety analysis allows organisations to track leading and lagging indicators that provide insight into the achieved level of safety and risk.
- **Organisational resilience** – Resilience is the ability of a system to absorb the unforeseeable. At an organisational layer, emergency (or crisis) response plans and

business contingency (or continuity) planning frameworks provide mechanisms to ensure that system management and organisational management actions are appropriate in response to foreseen or unforeseen events.

Plans should exist at many layers within an organisation to address individual system issues as well as larger organisation events. Resilience planning should be closely linked with risk management activities to ensure that hazards and hazard categories identified can be managed through resilience plans should other controls be ineffective.

- **Contingency planning** – A key element of resilience is organisational planning for contingency (or continuity arrangements) where systems or services are disrupted. Contingency planning is focused on return to system operation, prioritising essential systems and services. These may be restored in alternative forms initially.
- **Change management** – See the design-time definition. Any changes that occur during operation should be assessed for their safety impact and the associated safety risks should be identified and managed as part of change management.
- **Dynamic risk management** – The risk of systems in operation varies due to changes in the system and the operating environment. Risk levels are unlikely to stay constant and a dynamic approach to risk management is needed in response. Changes in the system that are not captured by change management activities should be considered on a regular basis along with changes in the operating environment. The effectiveness of controls should be regularly reassessed, as well as the threats that are posed during the systems operation. In-service safety performance monitoring should form a strong part of dynamic risk management alongside knowledge and experience of system stakeholders.
- **Digital twins** – A digital twin is a digital replica of a physical system. The ability to simulate real-world activities and integrate this analysis with data from actual experience provides a sophisticated approach to understanding impacts in operation before or even after occurrences.
- **Competency management/ staff training** – Procedures to ensure competency is maintained during system operation are a fundamental component to maintaining an assured operation. Competency must be maintained through management practices and training to ensure that changes in the system or operating environment are responded to, as well as the potential for degradation in human capability.
- **Diversity and inclusion in management** – As part of a strong safety culture, during operations it is important to embed diverse thinking into operations management, safety performance analysis and resilience functions.
- **Supply chain management** – Ensuring a good understanding of supply chain dependencies during operations and appropriate controls to mitigate issues in supply is a critical part of maintaining an organisation's safety performance in normal conditions and abnormal scenarios.
- **Safety/assurance cases** – During operations the logical safety or assurance case for operations must be maintained to ensure current management have confidence that the operation is acceptably safe to continue. The in-service safety or assurance case may vary in form from that used to introduce a system or change, as it will focus heavily on operational risk management and safety performance monitoring, as opposed to safety analysis and testing.

Operation-time controls for managing complexity safely at the task and technical layer include:

- **Self-monitoring** – The ability for systems to self-monitor their performance is an early way to gain insight into operational situations that may indicate failure and/or safety issues now or in the future.
- **Adaptation (optimisation)** – Adaptation allows a system to change in response to changes in the operating environment or the system itself. Adaptation can contribute to maintaining an acceptable operational safety performance.
- **Self-repair** – In combination with self-monitoring, self-repair allows systems to ensure an acceptable operational safety performance to be maintained or reintroduced following failure.
- **Run-time assurance** – Run-time assurance provides confidence that a system is operating as expected during operations. Real-time system monitoring in combination with a plan to be executed when failure occurs enhances the robustness of a system.
- **Human oversight** – The abilities of humans to interpret and act on information means that in some situations human oversight can play a critical role in managing system failures or operating environment changes. However, human oversight is not effective in all scenarios.
- **Cyber-security management** – Cyber threats pose significant

risks in modern complex systems and a mature cyber-security risk management approach is a critical element of preparing for such scenarios.

- **Task analysis** – Task analysis enables an understanding of how humans conduct activities as part of the system. This understanding is important as actions can vary between people, over time or in different environments. It is important to understand variance in task performance to maintain system performance.
- **Contingency rehearsals** – Practising contingency arrangements for system failure is key to ensure that arrangements are appropriate and that the capability exists to implement the arrangements on demand.

### A.3.6 Exacerbating factors

There are several factors that can make the management and governance of complex systems more challenging. They are viewed as 'exacerbating factors'. Exacerbating factors at the governance layer include:

- **No coordinating authority** – There is no regulatory or similar authority to ensure coordination of approaches to regulation and standardisation for example on an appropriate scale (which will often be global). For example, there is an absence of global structures for addressing AI and autonomous systems (although it should be noted that there are many initiatives in this space, such as [147]).
- **Opacity of decision-making** – Decisions are not visible to relevant/affected stakeholders so it is not possible to evaluate the decisions or to determine when circumstances arise such that those decisions are no longer appropriate, for example some of the decisions taken by the FAA regarding the Boeing 737 MAX were not visible even to other national airworthiness authorities until they were revealed by the Congressional enquiry [136].
- **Politicisation of decision-making** – Decisions driven by political (national or company-level) objectives and priorities rather than (scientific) evidence, such as differences in attitudes to the use of so-called high-risk-vendors in 5G networks in different countries [148], and decisions regarding the acceptance of Chinese-made masks for use by clinical staff dealing with COVID-19, see section 5.3.
- **Globalisation** – Products and systems can be made and used almost anywhere on the globe making it hard to track provenance, deal with known defects and recalls, and remedy environmental impact. An example of this is the PFAS chemicals used during the second half of the 20th century (see Appendix C.4.2).

Some of these factors interact, for example opacity of decision-making could arise, in part, from politicisation. A general concern is that, in some societies, the level of trust placed in 'experts' has been diminishing – so that we now have a so-called 'post truth era' with a lot of 'fake news' and 'alternative facts' [149], such as has happened with 5G and COVID-19 [150] where it has been said that 5G masts can contribute to the transmission of COVID-19.

Exacerbating factors at the management layer include:

- **Casualisation of labour/'gig economy'** – Increasingly people working on development or operation of systems are not permanent employees but are casual employees, making it hard to develop a safety culture and ensure learning from experience. An example of this is the 'safety driver' in the Uber Tempe accident (see Appendix C.2.2).
- **(Lack of) control of AI and autonomous systems** – Organisations used to managing conventional systems do not have the data collection, reporting and analysis capabilities necessary to manage AI and autonomous operations, for example the inability to assess 'ground truth' or internal system decisions when investigating mis-classification of objects by a perception system of an autonomous vehicle. Furthermore, due to path dependency within the deployed systems themselves, system instances may develop diverging behaviour in the field, making it impossible to make general statements about a fleet of systems.
- **Lean organisations** – Organisations 'shed' excess capacity or capabilities, preserving the minimum to enable them to carry out the business operations. This makes them less resilient, for example operating a 'zero stock' supply chain means that it is difficult to 'ramp up' supply if needed to respond to an emergency [151], or maintain production in the face of natural disasters [152].
- **Organisational memory** – The organisation fails to retain vital long-term knowledge, meaning that sustaining products and systems, or designing similar new ones is difficult, and this is often most difficult with tacit knowledge. This is especially problematic with systems produced in low volumes where there might be a long period between design activities, submarines and manned spacecraft.

Note that some of these factors interact, as poor organisational memory can be compounded by casualisation of labour as people who made the original design decisions or carried out safety analysis are not employed by the company, and may not even be

readily re-engaged if decisions need to be re-evaluated.

Exacerbating factors at the task and technical layer include:

- **Uncertainty** – Complexity can manifest itself as uncertainty in that it is impossible to place tight bounds on the range of possible system behaviours. System operators are faced with making decisions without clarity on how the system will react to their inputs, for example operator inputs contributed to some of the accidents of the Watchkeeper unmanned air system due to uncertainty in assessing the aircraft's state [134].
- **Disruptive technologies** – Most successful engineering design involves evolution of an existing successful design, [153] however the introduction of disruptive technologies can undermine the basis for the design leading to unanticipated failures. An example of this was problems with the introduction of composites in Rolls-Royce aircraft engines about 50 years ago [154], although the material is now being used on new large turbo-fan engines, suggesting that Petroski's model of learning from failure is still valid [155].
- **Long-tail dependencies** – Random variables that appear to exhibit no correlation can show long tail dependence in extreme situations, for example the correlated disruption of business, travel, and sport as a consequence of COVID-19 (see Appendix C.3.2).
- **Improbable events** – Low-probability high-impact events, especially those that can act as common-cause failures for otherwise independent systems or system parts, such as tsunamis or large-scale loss of communications capabilities such as the nine-hour loss of AT&T services that occurred in the US in 1990. This impacted many businesses as well as private individuals [156] (the US carrier network remains unreliable, but this is one of the largest single outage events due to its duration).

These factors do not seem to be either causes or consequences of the defining characteristics of complex systems, but they are factors that make management and governance more challenging and are often characteristics of systems of systems.

# B

# Stakeholder engagements

This section of the report summarises the feedback received through various stakeholder engagement activities.

# B.1 Stakeholder workshop

The study team conducted a stakeholder workshop at the Royal Academy of Engineering on 14 February 2020. The material here briefly summarises the output of the workshop and insights gained through wider stakeholder engagement. The wider engagement includes both discussions with particular individuals and an analysis of the results of an online questionnaire, with results summarised in Appendix B.2.

There were 24 invited attendees at the workshop, spanning academia, industry and regulators, and covering several domains including aerospace, automotive, built environment, defence, healthcare and retail services.

The summary highlights what seem to the study team to be the most significant points made during the day, however more detailed records from the workshop have been retained and will be used to shape the future conduct of the study, as appropriate.

### B.1.1 Characteristics of complexity

Complexity can arise from the system itself, from the system's environment, or from the interaction of the two. Some apparently simple systems, such as white goods, are complex because of the social and commercial context in which they are used, including other pressures such as 'net zero' and the circular economy. For example, reduced cost for electricity at night will encourage overnight charging of batteries, but with the attendant safety risk that this will not be supervised. There was a general consensus that the interconnectivity and interdependence of system components (and systems in a system-of-systems), was the biggest driver of complexity of the system itself. Growing interconnectivity is enabled by the Internet of Things (IoT) and this supports the development of 'open' systems, where boundaries change over time, and potentially rapidly.

Complexity arises from systems having multiple stakeholders, with each having different and potentially conflicting objectives. Often it is not clear where responsibility lies in the safe operation of such systems, for example home automation functions such as heating control or door/window unlocking using products and services from multiple vendors. Further, management of complexity can be exacerbated by different views of the scope and purpose of the system between stakeholders.

Similarly, complexity stems from a multiplicity of regulators, each having different and potentially conflicting rules and regulations. For example, in the case of a fire on a virtual bridge (a system enabling remote monitoring and operation of a vessel) in an onshore building, an evacuation is mandatory, which is in contradiction of the traditional duties of a ship captain.

There can be risk transference between stakeholders in a system, for example from the design and build phase of a building to facilities management, without awareness that this is happening, and hence no understanding of the need to manage those risks. An exacerbating factor is the dynamic nature of risk that can arise from changes in the system and from the environment.

There can be significant complexity in human-system interaction, and factors that can increase safety risk include high consequence decisions, speed of decisions, information overload, and loss of situational awareness; some of these factors have contributed to losses of unmanned air vehicles, such as Watchkeeper.

Systems suffer from 'path dependency' in that they are a product of, and they are constrained by, their history and how they have evolved. This can include technical evolution, operational changes but also the cultural 'baggage' of the environment that created them such as the practices and beliefs of the companies that produce certain products, as well as 'legacy' components.

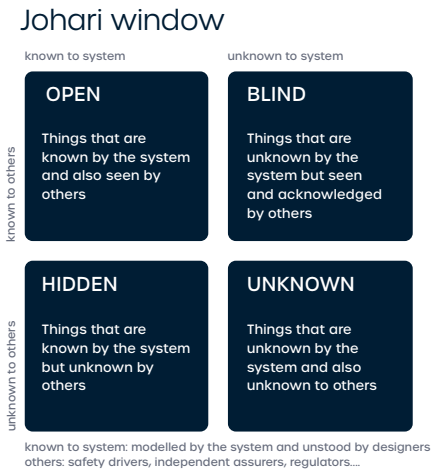



Figure 18: Johari window adapted to complex systems

It may be more helpful to think in terms of uncertainty than complexity, particularly uncertainty in decision-making, and it was suggested that the Johari window (initially developed in psychology) might usefully be adapted to complex systems. There are very different risk perceptions between different stakeholders, and these are strongly shaped by culture, context and social media; this is reflected to an extent in the Johari window. Finally, cyber-security and the potential for malicious behaviour increases complexity of safety management, even if not of the systems themselves.

### B.1.2 Safety management

The discussion of safety management focused almost entirely on operations, although it was noted that design needed to enable (safe) operations and that the transition from design and development to operations, for example facilities management for the built environment, needed to be carefully planned.

The five principles of high reliability organisations (HROs) are seen as key to managing safety of complex systems: preoccupation with failure, reluctance to simplify, sensitivity to operations, commitment to resiliency and deference to expertise. Related to the concept of HROs is that of 'mindful organisations' that are continuously looking for the unexpected.

Resilience and emergency response processes, including for evacuating buildings, have a role in safety management of complex systems. Principles such as rehearsals for incidents and the use of digital environments and immersive systems can be helpful as there is already a 'team understanding' when facing a real incident.

It is likely that outcome-based regulation will be most effective, and approaches such as the permissioning regimes used by the Health and Safety Executive (HSE) are likely to be effective and flexible, for example the HSE has recently extended permissioning to cover cyber-security issues.

It is important to monitor systems to look for leading indicators of problems, including monitoring of assumptions, although it is recognised that this becomes more difficult as complexity grows. There is a role for automation and data analysis here and monitoring mechanisms should also link to alerting systems. Another suggestion is that human instinct should not be ignored as it might often be a good predictor of problems (see the HRO concept of deference to expertise).It was also suggested that systems should not be built that exceed our ability to manage them safely, however judging this boundary is difficult due to the lack of an adequate theory of complexity.

'Near misses' (incidents) should be analysed as well as accidents to understand the causes and to make systems more resilient. It is desirable to take a 'systems approach' and to carry out the analysis in a 'no blame' or 'just' safety culture to encourage reporting and the most effective learning from experience. Other aspects of a safety culture including being proactive and empowering individuals to take 'ownership' of, and act on, safety issues are important too.

Those who will be exposed to risk from the system should be involved in its design and setting up safety management systems, and this should help combat probative blindness. In doing this it is appropriate to adopt principles of inclusive design, ensuring that different groups who may be affected by the system are properly represented. A particular suggestion was to 'design for the ageing' on the basis that doing this, rather than addressing the 'average', would produce systems usable by the majority of stakeholders.

It was widely recognised that safety management for complex systems require agility. Agile safety management can be done, and useful guidelines included: automating processes as well as the product, for example automate testing, use cloud-based deployment (with automatic roll-back) for IT-based solutions, don't separate development and support, and inform research from operations.

There is a 'lag' in developing standards and, as a consequence, in some areas industry is effectively making decisions on acceptable levels of risk. Agility therefore should also extend to governance and regulation, and the use of 'regulatory sandpits' providing a safe space for experimentation with new regulatory approaches was recognised as helpful in enabling regulators to understand how to deal with the challenges of growing system complexity.

It was also suggested that 'peer review' or even 'peer regulation' might be a useful alternative (or adjunct) to formal regulation, noting the difficulty of finding enough suitably qualified staff to work for regulators.

It was also noted that it is important to consider other factors, including economics and insurance, when considering the management of safety. There is also a need to consider global variations in economics and the value of a life in different countries, and how this affects risk acceptance.

Finally, it was noted that safety management has to be a 'whole lifecycle' consideration, and systems need to be designed to enable their management and governance, including updates, phasing out or replacing parts of systems, and eventual removal from service.

### B.1.3 Examples of complex systems

During the discussions many examples of complex systems were used to illustrate the ideas, and several accidents and incidents were identified for future investigation by the study. A broad range of examples, including those identified during the workshop, are described in Appendix C; several of the case studies are described using the framework introduced in this report.

### B.1.4 Observations

There are many drivers of system complexity. It would be easy to

focus on the system itself, but it was clear from the discussions and examples that it is important to recognise the environment of the system and human system interactions as sources of complexity.

As noted above, the discussions of safety management approaches referred almost exclusively to the operational phase of the system rather than the design phase. This demonstrates the need to manage the emerging safety properties of complex systems that cannot be (fully) predicted at design time and the need to continuously learn from experience. There is a challenge though of ensuring that the system is safe enough to be deployed in the field in the first place so that the experience can be gathered – and this is one issue that the Safer Complex Systems programme will need to explore.

Interconnectivity and interdependence, multiple stakeholders and human-system interaction were identified as some of the biggest drivers of complexity. Complexity results in a high level of uncertainty in the effectiveness of current safety management techniques. The stakeholders agreed that more emphasis must be placed on operational measures to manage emerging safety properties that cannot be predicted during design.

# B.2 Analysis of the questionnaire feedback

During the study a questionnaire was circulated to those who attended the workshops for the project; it was also circulated online and through various professional networks. Some of the findings from the questionnaire are summarised below.

Figure 19 shows the areas of the globe that respondents originate from. The majority of respondents are from the UK (28), however most global regions are represented at least once. Future studies could target questionnaires to get more coverage.

Figure 20 shows the different industry sectors the respondents work in. Ten respondents reported being in the transport sector (seven in civil aviation and rail, and three in transport equipment), five in defence, five in education, four in engineering.

Figure 19 shows that the majority of respondents (30) report that their sector is subject to some sort of formal regulation. A further two report partial formal regulation. A total of nine respondents report that their sector is not formally regulated, with one of those reporting that self-regulation is operating.

Figure 22 shows that there are diverse factors that cause complexity in the different sectors. Increasing automation and computerisation (nine), software/ technology (five), and artificial intelligence and machine learning (three) show a general theme of technology contributing to complexity. Regulation is also frequently reported (seven), along with other socio-political (four) and human factors such as people (six), expertise/skill (six) and market (three).

Breakdown of global regions where survey respondents work

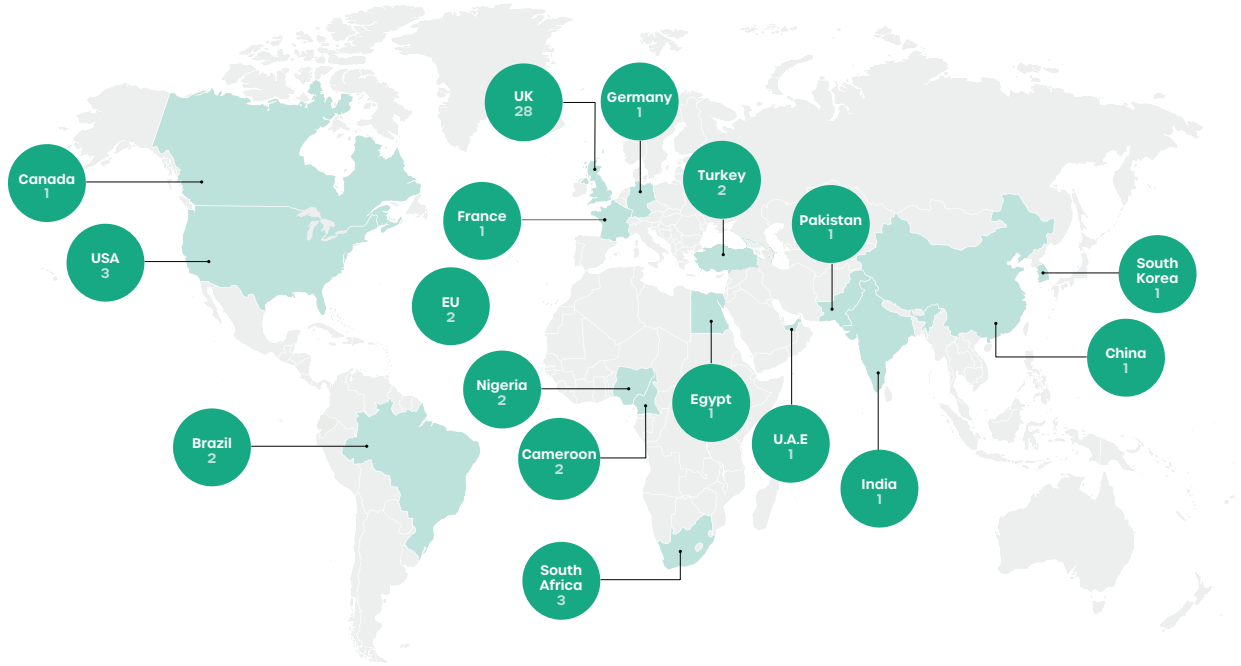



Figure 19 - Breakdown of global regions where survey respondents work.

## Sectors of work captured in our survey sample


Figure 20

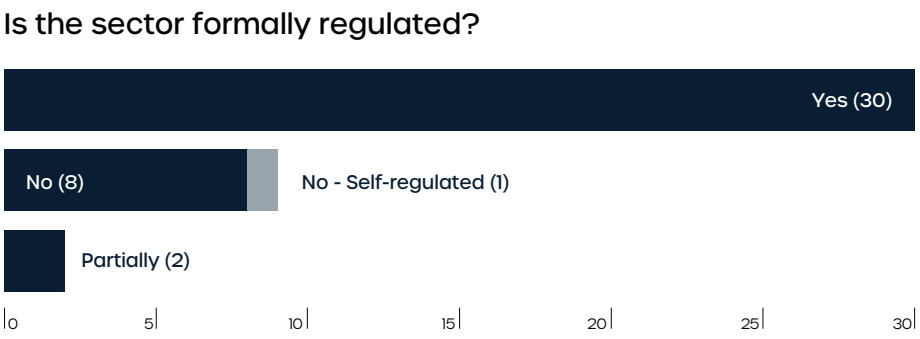



Figure 21 - Do people consider their sector of work to be formally regulated?

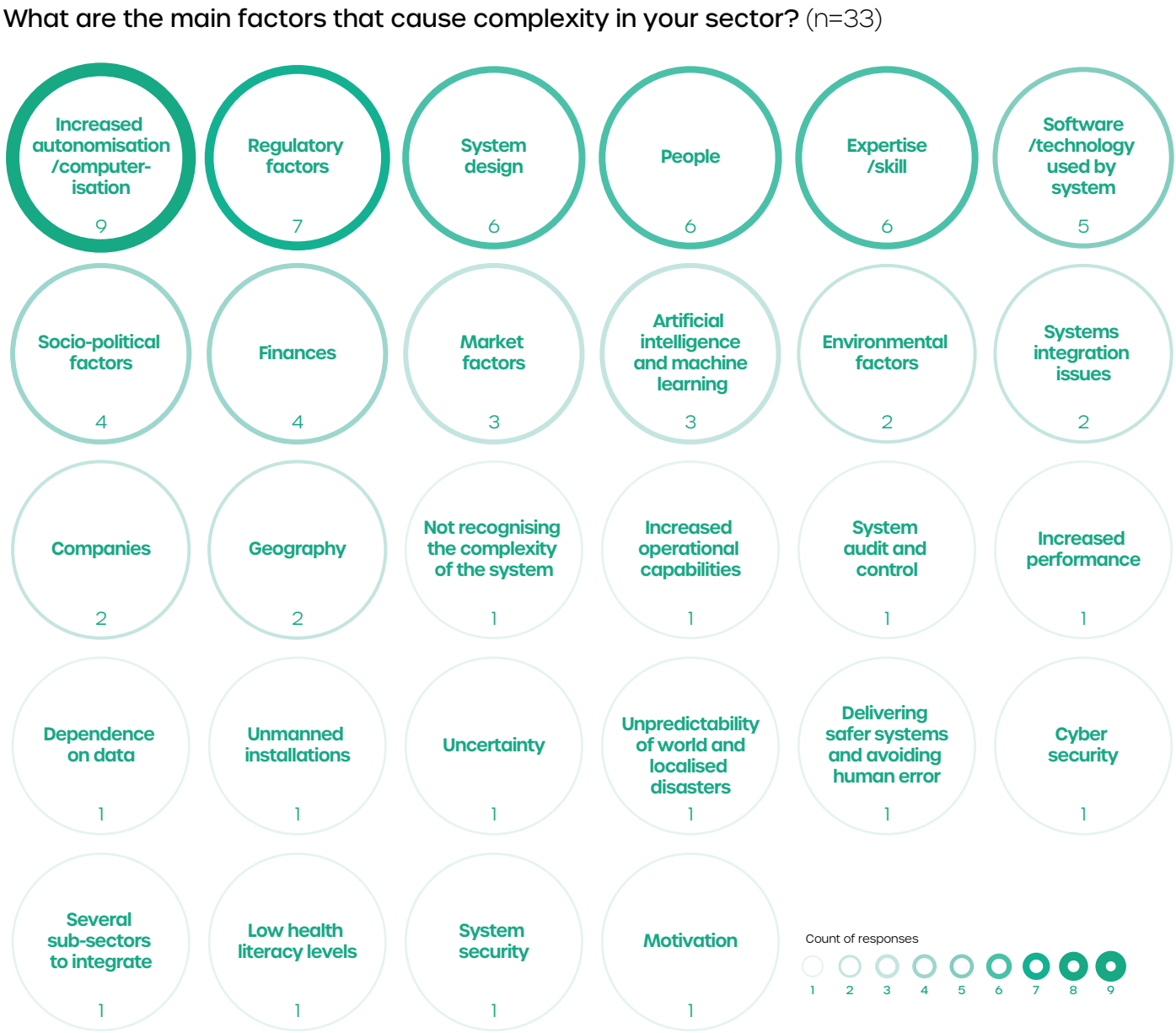



Figure 22 - Factors that contribute to the complexity in systems.

C

# Case studies considered during the study

This section of the report summarises the feedback received through various stakeholder engagement activities.

# C.1 Case studies from aeronautics and astronautics domains

### C.1.1 NATS system failure 12 December 2014

This case study is based upon [157] and is presented as a positive example of how planned operating-time strategies can be successful in delivering safe outcomes during occurrences in highly complex systems.

*What happened?*

On 12 December 2014 there was a failure at 2.44pm UTC of a computer system used to provide information to air traffic controllers managing the traffic flying at high level over England and Wales. This traffic includes aircraft arriving and departing from London airports as well as aircraft transiting UK airspace. The controllers put agreed procedures into action so as to limit traffic entering their area of responsibility and adopted manual methods for decision-making to ensure aircraft continue to maintain safe separation.

At 2.55pm all departures were stopped from London airports and at 3.00pm all departures were stopped from European airports that were planned to route through affected UK airspace. The computer system was restored to the controllers at 3.49pm, but without its normal level of redundancy. By 7.00pm, the engineering staff believed that they understood the cause of failure and full redundancy of the computer systems was restored at 8.10pm. Traffic restrictions were gradually lifted from 3.55pm as confidence increased, and the final restriction was lifted at 8.30pm. The disruption caused by the restrictions affected some airlines, airports and passengers into the following day.

There were no safety events recorded within the impacted airspace during the period of fallback operations or during the recovery phase.

*Why did it happen?*

The failure occurred due to a latent software fault that was present from the 1990s in a software application of more than two million lines of code. Design of software of any significant complexity is difficult and it is unrealistic to expect that software faults will not be introduced in development.

*The lesson for future complex systems operation*

Given the impracticality of designing 'perfect systems' and the emergent nature of outcomes in complex systems, operational management must include an effective set of planned recovery preparedness measures that can be implemented to ensure that consequences of system failure are mitigated.

*Critical factors in the event response*

The independent inquiry into the event identified two critical factors that enabled a rapid fault detection and system restoration:

- The software supplier's engineers (locally and internationally) had secure real-time access to data logs and were able to contribute fully to the incident response.
- The operational and technical

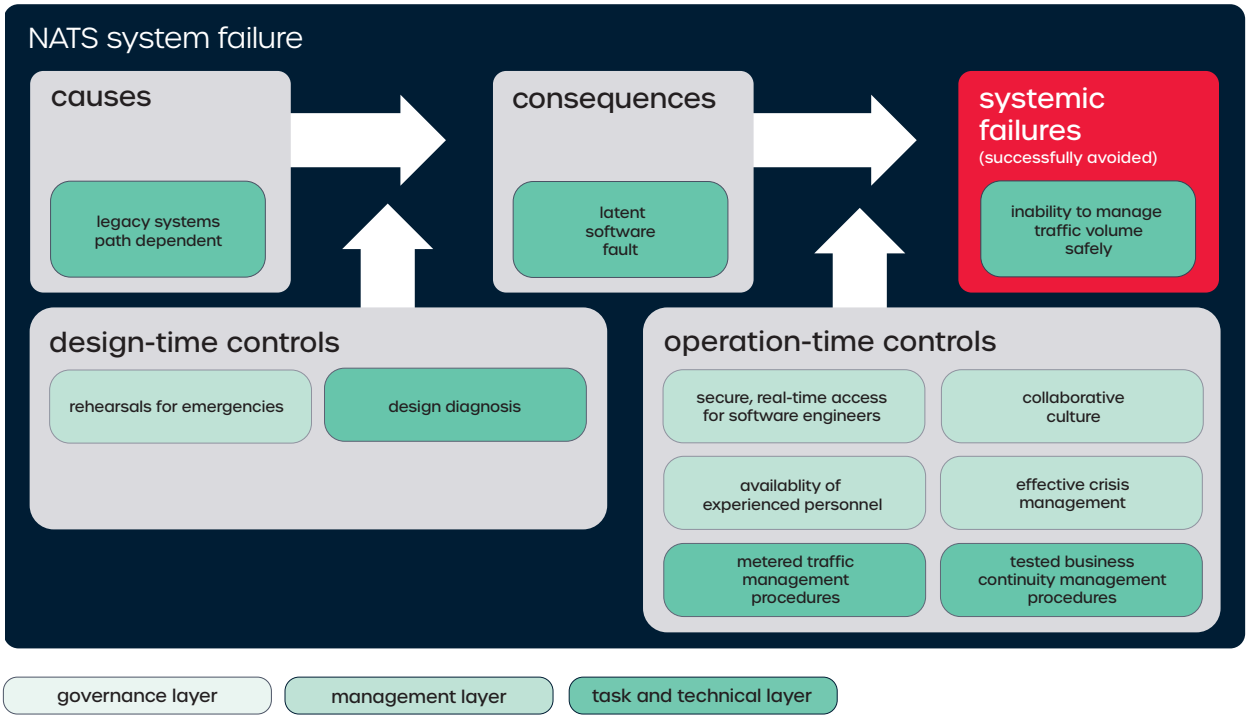



Figure 23: Analysis of the NATS system failure

teams exemplified a collaborative culture and their working was not hindered by organisational or commercial boundaries.

*Business continuity management good practices*

Several good practices were identified as part of the inquiry:

- In the event of a major system failure, the principles applied by NATS are to secure the safety of the operation through a reduction in traffic while considering the correct course of action to maintain the operation in a safe steady state, and then achieve full system and traffic recovery.
- NATS documents procedures for all anticipated failure modes via a series of fallback checklists and periodically exercise them.
- The timeliness of the response to the failure by NATS was impressive and comprehensive crisis management capabilities were mobilised quickly, including support from the contractor engineering design team, some of whom were based in the US, and therefore were in the middle of their normal working day.
- NATS had a well-established process that aimed to ensure staffing levels always meet routine roles and supervisory requirements in the direct conduct of their operations, and the Operational Resource Team was effective in ensuring that appropriate controller cover was in place throughout the recovery phase.
- Experienced senior controllers and technical experts were available in the Air Traffic Control Centre at the time of the event. Their presence, experience and expertise were viewed as key to the subsequent speed of analysis and decision-making in the operations room.

Resilience engineering and business continuity management planning form an integral part of the operational management of complex systems. While not all events can be prevented, management plans should prepare for foreseeable scenarios including preparing and empowering those likely to be involved.

Frameworks that empower operational staff and management to make risk-based judgements during crisis scenarios are required. Effective processes require the appropriate level of engagement and coordination of all stakeholders (including suppliers and customers) to achieve the best outcome. Consideration of appropriate culture and behaviours in and across organisations during a crisis event is an important part of preparation and planning.

### C.1.2 Boeing 737 MAX

This case study is heavily based upon [158] House Committee on Transportation and Infrastructure Preliminary Investigation Findings. It is presented as an example of how safety-critical issues in the complex system of aircraft manufacture and operation can manifest in terms of consequences of system complexity at the governance, management and task and technical layers. In this example, these consequences combined to cause catastrophic outcomes. It should be noted that these events remain under investigation and legal actions are outstanding: no conclusions are drawn in this report.

*What happened?*

On 29 October 2018, Indonesian carrier Lion Air operating flight 610 crashed into the Java Sea 13 minutes after take-off, killing all 189 passengers and crew. Less than five months later, on 10 March 2019, in strikingly similar circumstances, Ethiopian Airlines flight 302 – another 737 MAX aircraft – crashed six minutes after take-off, killing all 157 passengers and crew.

The 737 MAX was the 12th derivative model of the 737 aircraft (which was first certified in 1967) and the successor to the company's 737 Next Generation (NG) aircraft.

The US House Committee on Transportation and Infrastructure launched an investigation into the design, development and certification of the 737 MAX aircraft and related matters. The Committee held hearings on issues related to the 737 MAX; received an estimated 600,000 pages of records from Boeing, the FAA, airlines, and others; conducted 20 official interviews with current Boeing employees and FAA officials; and spoke with a wide range of aviation experts, engineers, software development experts, and former FAA and Boeing employees. The information below is based upon the Committee's preliminary findings.

*Why did it happen?*

According to the Committee's preliminary report, while multiple factors led to these accidents, both crashes shared a key contributing factor: a new software system called the Maneuvering Characteristics Augmentation System (MCAS), which Boeing developed to address stability issues in certain flight conditions induced by the plane's new, larger engines, and their relative placement on the 737 MAX aircraft compared to the engine placement on the 737 NG.

The Committee's preliminary findings identify five central themes that affected the design, development and certification of the 737 MAX and FAA's oversight of Boeing. Acts, omissions, and errors occurred across multiple stages and areas of the development and certification of the 737 MAX.

- **Production pressures:** According to the Committee's preliminary report, there was tremendous financial pressure on Boeing and subsequently the 737 MAX programme to compete with the

Airbus A320neo aircraft. Among other things, this pressure resulted in extensive efforts to cut costs, maintain the 737 MAX programme schedule, and not slow down the 737 MAX production line. The Committee's investigation identified several instances where the desire to meet these goals and expectations jeopardised the safety of the flying public.

- **Faulty assumptions:** According to the Committee's preliminary report, Boeing made fundamentally faulty assumptions about critical technologies on the 737 MAX, most notably with MCAS. Based on incorrect assumptions, Boeing permitted MCAS software designed to automatically push the plane's nose down in certain conditions to rely on a single angle of attack (AOA) sensor for automatic activation, and assumed pilots, who were unaware of the system's existence in most cases, would be able to mitigate any malfunction. Partly based on those assumptions, Boeing failed to classify MCAS as a safety-critical system, which would have offered greater scrutiny during its certification. The operation of MCAS also violated Boeing's own internal design guidelines established during development.
- **Culture of concealment:** According to the Committee's preliminary report, in several critical instances, Boeing withheld crucial information from the FAA, its customers, and 737 MAX pilots. This included hiding the very existence of MCAS from 737 MAX pilots and failing to disclose that the AOA disagree alert was inoperable on the majority of the 737 MAX fleet, despite having been certified as a standard cockpit feature. This alert notified the crew if the aircraft's two AOA sensor readings disagreed, an event that occurs only when one is malfunctioning. Boeing also withheld knowledge that a pilot would need to diagnose and respond to a 'stabilizer runaway' condition caused by an erroneous MCAS activation in 10 seconds or less, or risk catastrophic consequences.
- **Conflicted representation:** According to the Committee's preliminary report, the Committee has found that the FAA's current oversight structure with respect to Boeing creates inherent conflicts of interest that have jeopardised the safety of the flying public. The Committee's investigation documented several instances where Boeing authorised representatives (ARs) (Boeing employees who are granted special permission to represent the interests of the FAA and to act on the agency's behalf in validating aircraft systems and designs' compliance with FAA requirement) failed to take appropriate actions to represent the interests of the FAA and to protect the flying public.
- **Boeing's influence over the FAA's oversight:** According to the Committee's preliminary report, multiple career FAA officials have documented examples to the Committee where FAA management overruled the determination of the FAA's own technical experts at the behest of Boeing. In these cases, FAA technical and safety experts determined that certain Boeing design approaches on its transport category aircraft were potentially unsafe and failed to comply with FAA regulation, only to have FAA management overrule them and side with Boeing instead.

### C.1.3 Urban Air Mobility

In this section we describe a future development of the aviation system, which provides a good example of how complexity will continue to increase (most likely at an ever increasing rate). Many of these types of challenges will exist in other parts of the aviation system including traditional regular public transport, high altitude transport and future supersonic transport.

Urban air mobility (UAM) is the name currently used for the transportation of people by air in an urban environment and was developed as a concept in response to urban transportation congestion. UAM is not a new concept and mature air taxi markets using helicopters exist in a few cities, such as Sao Paulo, Brazil.

The recent development of new types of aircraft including electrical vertical take-off and landing (EVTOL) vehicles and hybrid-electric vehicles has led to significant investment in the development of a future mass-market UAM industry. Expectations are that commercial air taxi services using new vehicle types will commence in the next five years. In the US, predictions have been made that the market has a potential demand of 55,000 daily trips with 85,000 passengers using 4,000 aircraft [159]. The associated financial estimate is that this will lead to an initial annual market value of $2.5 billion rising to up to $500 billion in the US alone.

At the same time, significant investment is occurring in developing unmanned aircraft systems (UAS) for a range of applications including cargo or package delivery and a variety of survey activities.

Given the costs of aircrew, the business case for UAM is potentially made stronger with the use of autonomy where pilotless EVTOLs are used. Initial UAM operations are expected to be piloted, however industry participants are planning for future autonomous UAM operations. A similar industry maturity roadmap is expected for UAS and it is expected that traditional aviation will see increasing use of autonomy over the same period.

Some key challenges related to the complexity of a future UAM system are outlined below and provide examples of how complexity is predicted to increase significantly in systems soon.

- **Safety performance:** Current regular public transport (RPT) aviation is a good example of a well-managed system which achieves one of the highest safety performances in the transportation industry. The public acceptance of a new mass-market UAM industry will be contingent on delivering a safety performance which is socially accepted. Establishment of safety frameworks to deliver outcomes at an appropriate level will require new thinking. Ensuring that the new UAM industry does not have competency gaps to deliver an appropriate safety performance will be a key challenge at the governance and management layers. Ensuring these safety objectives are not in competition with delivering new services will also be important.
- **Public acceptance:** As well as safety outcomes for passengers and those below in the urban environment, public support for UAM will be contingent on other issues being well managed, including privacy and noise. Issues of safety, privacy and noise may be inter-related and will be difficult to manage in isolation. The voicing of these issues by the community may be inter-related, where safety is used as a proxy for noise or privacy concerns. Ensuring a clear understanding of how the jurisdictions inter-relate at the governance layer will be important to ensure societal concerns are appropriately managed.
- **Interconnectedness:** The level of interconnectedness of the aviation system with other industries will increase significantly through the introduction of UAM. Examples of areas of greater interconnectedness include urban planning, urban transportation systems, energy/power infrastructure, telecommunications infrastructure.
- **Rapid technology change:** UAM implementation is contingent on the use of new technologies and production methods not previously used in aviation (for example mass production of vehicles, use of different materials and systems). The rapid technology changes that are occurring will need to be successfully managed by the governance layer in aviation, which is historically slow to evolve. The potential for competency gaps and standards lag must be mitigated.
- **Autonomy:** The introduction of autonomy will be challenging in its own right, however autonomous systems will be required to integrate with traditional piloted aircraft with voice-based communications. The management of autonomous systems as well as the technical assurance of these systems will be potential exacerbating factors to be mitigated.

### C.1.4 NASA Challenger disaster

The flight of the Space Shuttle Challenger on Mission 51-L began at 11.38am EST on 28 January 1986 [160]. It ended 73 seconds later in an explosive burn of hydrogen and oxygen propellants that destroyed the external tank and exposed the Orbiter to severe aerodynamic loads that caused complete structural breakup. All seven crew members died.

The consensus of the Commission and participating investigative agencies was that the loss of the Space Shuttle Challenger was caused by a failure in the joint between the two lower segments of the right solid rocket motor. The specific failure was the destruction of the seals that are intended to prevent hot gases from leaking through the joint during the propellant burn of the rocket motor.

The decision to launch the Challenger was flawed. Those who made that decision were unaware of the recent history of problems concerning the O-rings and the joint and were unaware of the initial written recommendation of the contractor advising against the launch at temperatures below 53 degrees Fahrenheit and the continuing opposition of the engineers at Thiokol after the management reversed its position. They did not have a clear understanding of Rockwell's concern that it was not safe to launch because of ice on the pad. If the decision-makers had known all of the facts, it is highly unlikely that they would have decided to launch 51-L on 28 January 1986.

### C.1.5 Air France Flight 296

On 26 June 1988 at Mulhouse-Habsheim Airport an Airbus A320, registration FGFKC, crashed 300 metres beyond the end of runway. Onboard were two flight crew, four cabin attendants and 130 Passengers.

As summarised in the accident report [161], as part of an airshow, the aircraft flew over runway 34R at a height of approximately 30 feet, engines at idle, with an angle of attack increasing up to the maximum possible taking into account the deceleration rate of the aircraft. During the go-around, the aircraft touched the trees a short way beyond the end of the runway, sank into the forest, came to rest and caught fire. Evacuation was undertaken immediately but three passengers died in the fire.

In the accident report the Commission determined that the accident resulted from the combination of the following conditions:

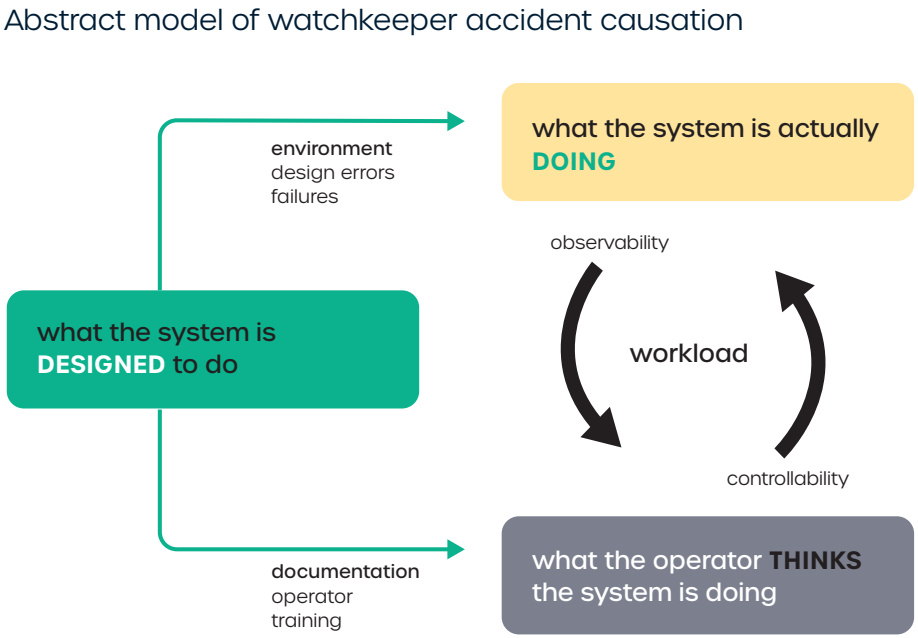



Figure 24: Abstract model of Watchkeeper accident causation

- very low flyover height, lower than surrounding obstacles
- speed very slow and reducing to reach maximum possible angle of attack
- engine speed at flight idle
- late application of go-around power.

This combination of conditions led to impact of the aircraft with the trees. The Commission also believed that if the descent below 100 feet was not deliberate, it may have resulted from failure to take proper account of the visual and aural information intended to give the height of the aircraft. The Commission also identified several other factors that contributed towards placing the crew in a situation that they were not able to fully control.

### C.1.6 Überlingen mid-air collision

On 1 July 2002 at 9.35pm a collision between a Tupolev TU154M, which was on a flight from Moscow, Russia to Barcelona, Spain, and a Boeing B757-200, on a flight from Bergamo, Italy to Brussels, Belgium, occurred north of the city of Überlingen (Lake of Constance). Both aircraft flew according to instrument flight rules and were under control of ACC Zurich. After the collision both aircraft crashed into an area north of Überlingen. There were a total of 71 people on board of the two airplanes, none of whom survived the crash. The German Federal Bureau of Aircraft Accidents Investigation (BFU) identified the following immediate causes [162]:

- The imminent separation infringement was not noticed by air traffic controller (ATC) in time. The instruction for the TU154M to descend was given at a time when the prescribed separation from the B757-200 could not be ensured anymore.
- The TU154M crew followed the ATC instruction to descend and continued to do so even after Traffic Collision Avoidance System (TCAS) advised them to climb. This manoeuvre was performed contrary to the generated TCAS Resolution Advisory (RA).

The BFU identified the following systemic causes:

- The integration of Airborne Collision Avoidance System (ACAS)/TCAS II into the system aviation was insufficient and did not correspond in all points with the system philosophy.
- The regulations concerning ACAS/TCAS published by ICAO and as a result the regulations of national aviation authorities, operations and procedural instructions of the TCAS manufacturer and the operators were not standardised, were incomplete and partially contradictory.
- Management and quality assurance of the air navigation service company did not ensure that during the night all open workstations were continuously staffed by controllers.
- Management and quality assurance of the air navigation service company tolerated for years that during times of low traffic flow at night only one controller worked and the other one retired to rest.

### C.1.7 Watchkeeper accidents

The British Army uses an unmanned air system (UAS) known as Watchkeeper for reconnaissance. It is operated and supported by a ground crew at a ground control station (GCS) but some functions are autonomous, such as the ability to execute a go-around (rather than landing). Watchkeeper does not use AI or ML but it is a good example of problems that can arise when

sharing control between operators and a (semi-)autonomous system. Watchkeeper has suffered five accidents to date; a far higher loss rate than the safety analysis predicted [163]. The UK Defence Safety Authority (DSA) report on the fourth Watchkeeper crash [164] draws out three 'themes' from all five accidents (the fifth was being investigated at the time of writing that report). The themes highlighted are:

- The incomplete understanding of the full system, and of how sub-systems integrate.
- The need to improve collection and analysis of data.
- Ground crew and engineer workload.

For example, the report cites the high rate of warnings, cautions and advisory (WCA) notifications creating a high ground crew workload. Further, the ground crew rely on their collective knowledge to understand the situation and how to respond, for example to interpret WCAs rather than referring to documentation (paper or electronic). Based on analysis of the reports and from discussions with the manufacturers [163], the 'high-level' accident causation is as set out in Figure 24.

The design model used, including for safety analysis, was not an accurate predictor of the actual behaviour of the system (in its operational environment). The ground crew training and documentation did not help them to understand the actual system behaviour, including WCAs. Finally, workload had a significant impact on the operators' ability to observe the state of the system and to control it. There is a dissonance between the three sub-models in Figure 24, which suggests the distinctions made between 'work as imagined' and 'work as done' [73]. The Watchkeeper accidents were one of the initial motivations for developing the framework presented in this report.

The accidents suffered by the British Army's Watchkeeper system demonstrate the distinction between 'work as imagined' and 'work as done', illustrating issues at all levels of the framework presented in this report.

# C.2 Case studies from the mobility domain

### C.2.1 Smart motorways

*What happened?*

Smart motorways use adaptive speed limits and allow drivers to use the hard shoulder directed by electronic signs, both to reduce congestion and to close lanes to reduce the impact of accidents or breakdowns. The original safety case [165] in the UK was based on a trial that included emergency refuges for vehicles in trouble every 500 metres and a reduced speed limit in the hard shoulder lane. Evidence collected by the highways agency [93] during the trial period on a stretch of the M42 motorway indicated a 22% improvement in journey reliability and a reduction in personal injury accidents by half.

However, during roll-out of the system, modifications were made allowing refuges to be provided every 1.6 kilometres and allowed the usual motorway speed limit of 70 miles per hour (113 km/h). A stopped-vehicle detection system was also not put in place consistently across the network. The result of these changes was a significant increase in accidents in those stretches of road, which in turn eroded public trust in the system resulting in negative press coverage and the reaction of the government [166] to review the continuation of the scheme.

*Why did it happen?*

Due to limited information available, some of the causes of the problems associated with the deployment of this system can only be hypothesised and a complete analysis is therefore not possible. However, this case study highlights the impact that decisions at a governance and management level can have on the overall safety and eventual public acceptance of a new technology. Although the technology had been demonstrated to improve overall safety during trials, when deploying the system, the boundary conditions vital to ensuring safety were not considered (speed limits, closely spaced refuges, stopped-vehicle detection). This indicates that between the initial safety analysis and deployment, an additional analysis taking into account these dependencies and impact of context changes was not made or was ineffective. The case study also raises questions regarding drivers' interpretations of dynamic rules on the road and how these can be enforced (for example definition and enforcement of punishments for disregarding lane closed signs).

### C.2.2 Uber ATG Tempe, Arizona crash

*What happened?*

On the evening of 18 March 2018, an automated test vehicle operated by Uber Advanced Technologies Group (Uber ATG) in Tempe, Arizona was involved in an accident that fatally injured a 49-year-old pedestrian crossing a dual carriageway while pushing a bicycle. The circumstances surrounding this incident highlight many of the risks involved in introducing automated driving technologies at the technical, management and governance layers.

At the time of the accident, the vehicle was in automated driving mode, travelling within the speed limit of 45 miles an hour at night on a dry illuminated road.

The test vehicle was being operated by a safety driver, who at the time was watching entertainment content on a mobile phone, contrary to the company's operating procedures.

Approximately 5.6 seconds before the crash, the pedestrian was detected as an object by the vehicle. However, up until the impact, the vehicle variously misclassified the pedestrian as a vehicle, unknown object and bicycle. On each new classification, the object trajectory prediction algorithm would 'reset' and, without taking into account the previously observed trajectory, assign a new classification-dependent goal-based trajectory prediction. Not once did the system correctly identify the object as a pedestrian pushing a bicycle.

The system identified the imminent collision 1.2 seconds before impact, however in order to avoid the consequences of false-positive misclassifications, the system was designed to suppress any braking manoeuvres in such a case, in the assumption that an attentive operator would take control. In this case, due to distraction, the operator was not able to react in a timely enough manner to prevent impact. Furthermore, emergency braking systems pre-installed within the vehicle had been deactivated in order not to conflict with the prototypical functions under test.

Toxicological tests performed on the deceased showed traces of drugs that may have impaired her perception and decision-making capabilities. However, neither the decision to cross the street in a prohibited area nor the possible impairment through medication or drugs were mitigating factors of the vehicle not being able to detect and avoid the pedestrian. In fact, from an ethical perspective both factors would not have been able to be observed by the vehicle and should therefore not impact any judgements it makes as other legitimate factors could have caused the pedestrian to cross the street at that time.

*Why did it happen?*

The subsequent accident report by the US National Transportation Safety Board [167] assigned the inattentiveness of the operator resulting from 'automation complacency' as the most probable cause of the crash. However, it also identified several additional contributing factors, including inadequate safety

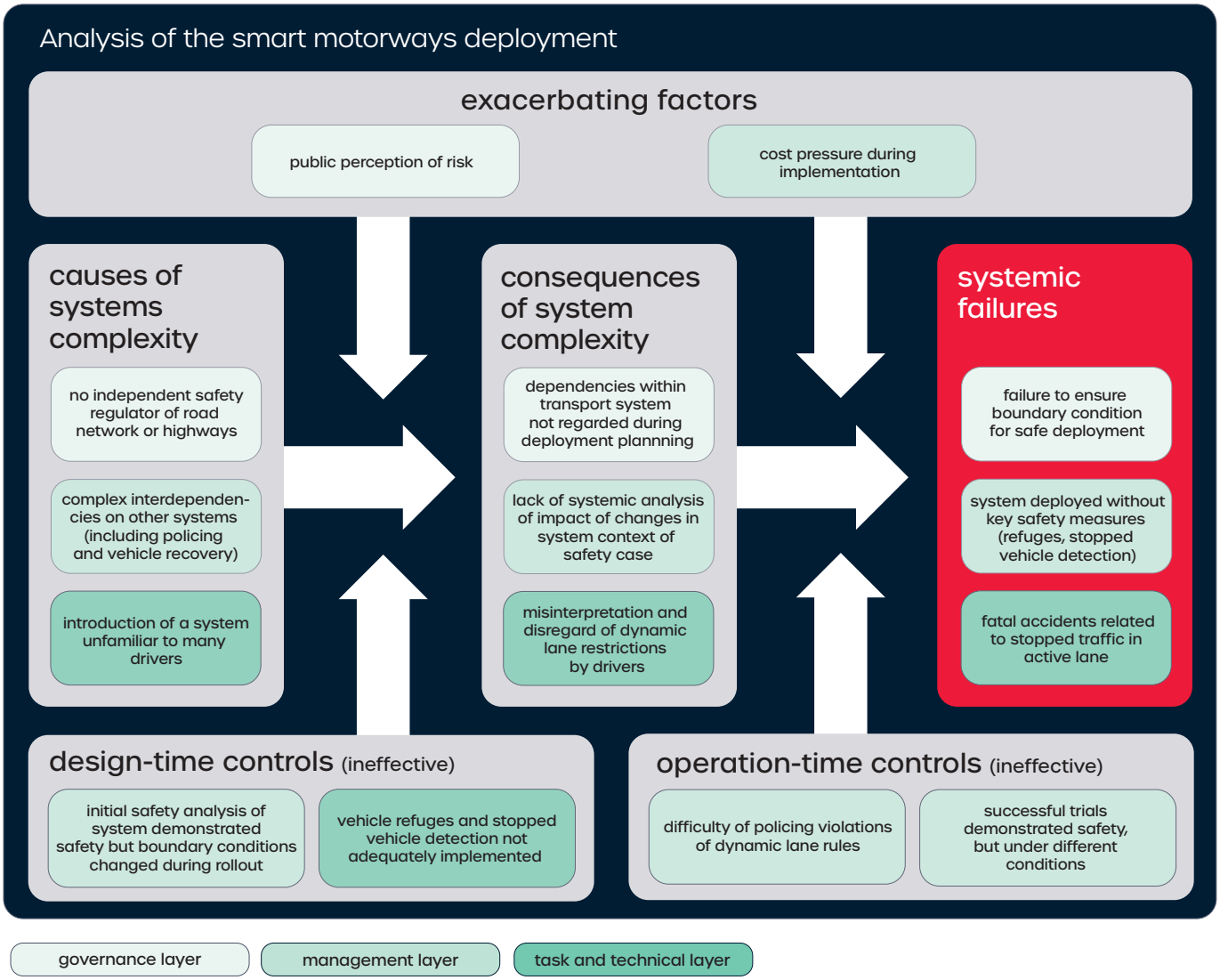



Figure 25: Analysis of the smart motorways deployment

risk assessment procedures at Uber ATG, and ineffective oversight of the vehicle operators, including lack of mechanisms for addressing operators' automation complacency. Additional mitigating factors were identified as the ambiguous nature of the ground separating the directions of the carriageway, which appeared to include pedestrian walkways, and ineffective oversight of automated vehicle testing by Arizona's Department of Transportation.

Automated driving, in itself, is an inherently complex task due to the unpredictable nature of the operating environment and road users, as well as the technical limitations of current technology. There is, as yet, no consensus regarding state-of-the-art techniques for developing and validating these systems to an adequate level of performance for complex urban environments. However, it is clear from an analysis of this accident that there was no need for the technical deficiencies of the system to lead to this fatality and that the safety culture within Uber ATG was a contributing factor, as well as the inadequate legislative constraints. The factors contributing to the accident are summarised in Figure 26, based on the structure of the framework presented in the primary deliverable of this study.

From this perspective, the consequences of functional insufficiencies of the system at the technical level can be seen as an emergent property of the management and governance levels and the inadequacy of the duty holders to understand and manage the risks associated with operating such systems. There were insufficient measures in place at the governance and management level to constrain the emergent risk of deploying the technical system in its environment. This includes the impact of automation complacency and lack of oversight regarding the behaviour of the operator, as well as the impact that road layout had on the decision by the pedestrian to cross the street at that point.

The case study also demonstrates the conflicting pressures to promote innovation in technologies such

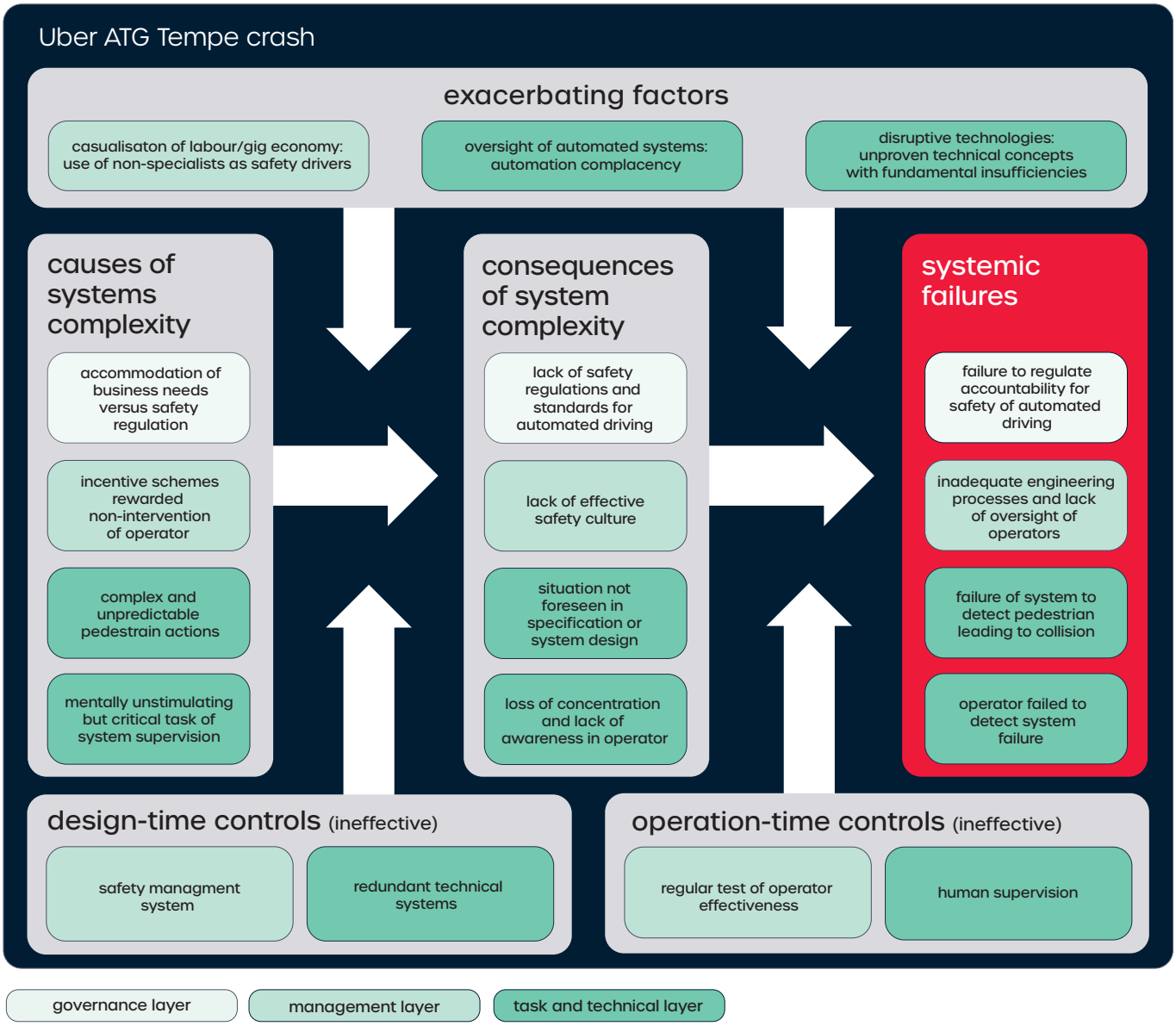



Figure 26: Analysis of the Uber ATG Tempe crash

as automated driving that have the potential for improving overall road safety, while in parallel managing the risk of integrating such technologies into complex environments with an insufficient understanding of emergent behaviours.

### C.2.3 Tesla autopilot crash

This incident [168] involved an autopilot feature failing to detect an articulated truck crossing laterally in front of the vehicle resulting in a fatal accident. The accident was caused by a combination of factors including the road layout and the truck's failure to yield right of way, automation complacency and misuse on the side of the Tesla driver, as well as a combination of technical insufficiencies in the autopilot and automated emergency braking systems of the vehicle itself. As such, underlying factors were very similar to the Uber ATG incident (see above) despite a lower level of automation in the vehicle.

### C.2.4 GM ignition switch problem

This case study [90] involves the ignition switch failing in a number of vehicles when the key fob was knocked by the driver's knee, subsequently deactivating power steering and causing the engine to stall while at the same time causing the airbags not to deploy, exacerbating the consequences of the accidents and resulting in fatalities. In trying to determine the cause of ignition-related issues in the vehicles, engineers were distracted by unrelated problems that were occurring at the same time with the ignition system causing the root cause of the failure to be overlooked. Aside from the ignition switch's technical problems, several organisational issues, including cost focus, structural secrecy, a lack of urgency, inadequate oversight and a company culture characterised by low accountability, contributed to the causes of the risks.

### C.2.5 Jeep Cherokee hack

There have been a number of examples of academic security researchers exploiting vulnerabilities in vehicle electronics and software architectures in order to remotely control safety-critical functions such as steering and braking. The most prominent of these was the Jeep Cherokee hack performed and documented in 2015 [169]. The researchers demonstrated that the particular architecture and implementation of this vehicle presented a large attack service for remote manipulation.

Increasing connectivity between automotive components themselves, including between infotainment and chassis control domains, coupled with a lack of system-level analysis of cyber-security related risks, led to vulnerabilities being missed in the design process. In addition, the use of commercially available and open source software components mean that any vulnerabilities left unpatched in software would be well-known to potential hackers.

High profile cases such as the Jeep Cherokee hack have encouraged the automotive industry to consider cyber-security as a core dependability property of vehicle architectures. However, the increasing connectivity of vehicles and the development of side-channel attacks [123] demonstrate that this is an area that is also sensitive to increasing complexity and therefore would also profit from the approach described within this study. This topic highlights the need for operation-time control measures as vulnerabilities must be continuously addressed, both as they are discovered and as the capabilities of hackers improve over the entire lifecycle of the vehicle.

# C.3 Case studies from the healthcare domain

*The views and opinions expressed in this case study are those of the University of York research team and do not necessarily reflect the views of Engineering X.*

Healthcare has an enormous scope and encompasses issues that affect a single individual up to global pandemics that have the potential to affect everyone on the planet. The aim in the two case studies here is to illustrate those extremes – the first example is treatment of a patient with sepsis, and the second is a discussion of (some aspects of) COVID-19.

### C.3.1 Sepsis fatality

This example highlights the fact that, in healthcare, we are often concerned about safety risks to specific individuals, not the risk to a population considered in most other domains.

#### *What happened?*

At 9.35am on 21 October 2012 a 31-year woman referred herself to the gynaecology ward at University College Hospital Galway, when she was 17 weeks pregnant. She complained of having had lower backache for the previous 12 hours. Analysis including examination of urine showed that the patient's condition was normal. At the patient's request the foetal heart rate was monitored and recorded. A course of pain management was recommended. The patient was reassured but told to come back if she had any concerns.

She returned to the hospital at 3.30pm saying she had "felt something coming down" and had "pushed a leg back in". The documentation said that there had been no vaginal fluid loss. Examination indicated bulging membranes and no cervix to be felt and the clinicians concluded that there was "an inevitable/impending pregnancy loss". She was admitted to the hospital for management of inevitable miscarriage on the same day.

At 00.30am on 22 October, the patient's membranes spontaneously ruptured. She continued to be monitored and given fluids and oral antibiotics. Fluids were later discontinued but oral antibiotics were given regularly. The clinicians continued to monitor the foetal heart rate. The patient asked about termination but a consultant stated: "Under Irish law, if there's no evidence of risk to the life of the mother, our hands are tied so long as there's a foetal heart". On 24 October the patient's condition deteriorated and she was diagnosed with sepsis and chorioamnionitis. An ultrasound scan at 3.00pm on 24 October showed that there was no movement of the foetal heart. She was transferred to the operating theatre to insert a central venous monitoring line and she had a miscarriage with spontaneous delivery. She was moved into the High Dependency Unit at 4.45pm on 24 October.

The patient continued to deteriorate with increased need for oxygen and vasopressors (part of standard sepsis treatment) and was transferred to the Intensive Care Unit (ICU) at 3.00am on 25 October. Despite intubation and mechanical ventilation, the patient's condition deteriorated further in the ICU and she died at 1.09am on 28 October 2012.

#### *Why did it happen?*

The subsequent investigation occurred under the auspices of the Health Service Executive (HSE) in Ireland. The approach followed the HSE's guidelines on Systems Analysis Investigation of Incidents and Complaints [170]. The subsequent report [171] found three key causal factors in the accident; these findings are presented here but restructured so as to show the mapping to the three layers in our framework: governance, management and task (there is no substantial technical element in this case).

In Ireland, the Constitution says: "The State acknowledges the right to life of the unborn and, with due regard to the equal right to life of the mother, guarantees in its laws to respect, and, as far as practicable, by its laws to defend and vindicate that right." The implications of this part of the Irish Constitution have been clarified through case law, and the guidance available to clinicians in the Guide to Professional Conduct and Ethics for Registered Medical Practitioners [172] produced by the Medical Council states: "Abortion is illegal in Ireland except where there is a real and substantial risk to the life (as distinct from the health) of the mother", reflecting the case law. In this case, uncertainty about the legal situation contributed to delays in treating the patient.

At the management level an underlying factor is that initially the doctors dealing with the patient were junior and under heavy workload. They were not always able to assess the patient quickly, but the case wasn't escalated for a senior doctor's review. They seemed to have been strongly influenced by the legal situation and the difficulty of interpreting the law when there is a potential major hazard to the mother's life. However, there is widespread understanding that a foetus under 24 weeks will not survive following delivery and it was recognised in the report [171] that international best practice is: "Expediting delivery at the earliest signs of infection in the uterus is a critical part of management to reduce the risk of progression to sepsis, severe sepsis and septic shock and maternal morbidity and death". Further, if the mother has sepsis the foetus will not survive anyway.

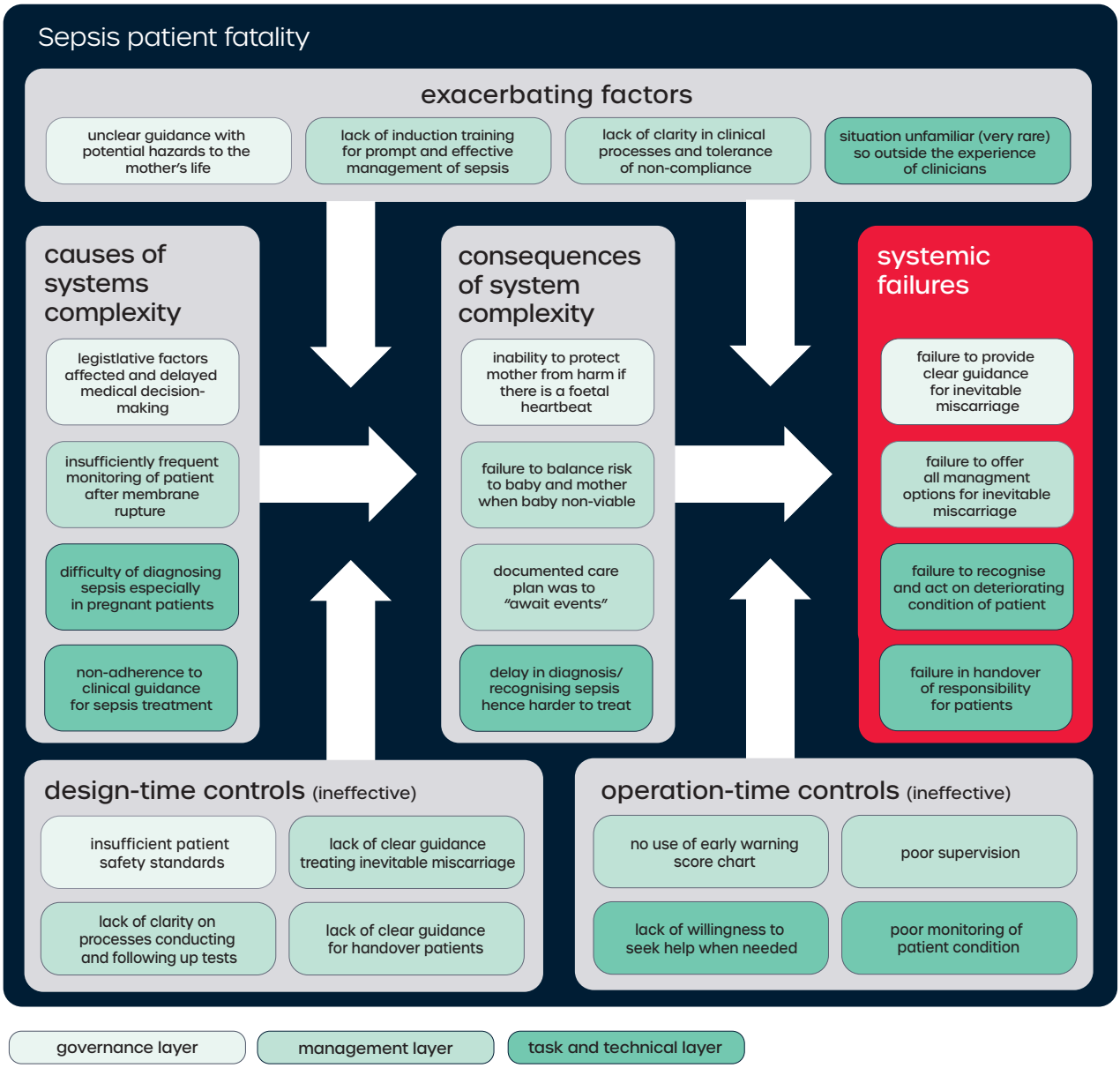



Figure 27: Analysis of factors in sepsis patient fatality

The lack of clarity in, or possibly lack of, processes and guidelines and the lack of training in the management of sepsis are clear management failings that have a major influence on the task-level issues.

Successful management of sepsis requires early diagnosis and treatment, and this did not happen. It should be noted, however, that the patient's condition is rare and contributed to the difficulty of (or lack of familiarity with) managing the situation. The natural physiological changes brought about by pregnancy make it harder to diagnose sepsis and the report stated that there were no national or international guidelines on inevitable miscarriage at this stage in pregnancy at that time.

The main task-related factor was the "inadequate assessment and monitoring to enable clinicians to recognise and respond to signs that the patient was deteriorating" [171]. This included recognising that the most likely cause of the inevitable miscarriage was infection that could escalate to sepsis, and the failure to develop an adequate care plan. Indeed, up to the morning of 24 October, the plan was to keep monitoring and "await events". The task-level analysis breaks this overall finding down into the main contributory causes which, as noted above, are heavily influenced by the management layer issues.

It is worth noting that the failures may seem obvious with hindsight, but they were not at the time; this is likely to be quite common in very busy and stressful activities such as frontline healthcare.

### C.3.2 Coronavirus What happened?

Coronaviruses are a family of viruses that include the common

| Country, other | Total cases | New cases | Total deaths | New deaths | Total recovered | Active cases | Serious, critical | Total cases / 1M population | Deaths / 1M population |
|---|---|---|---|---|---|---|---|---|---|
| WORLD | 841,378 | +56,719 | 41,403 | +3,365 | 176,443 | 623,532 | 31,565 | 107.9 | 5.3 |
| USA | 180,340 | +16,552 | 3,573 | +432 | 6241 | 170526 | 3981 | 545 | 11 |
| Italy | 105,792 | +4,053 | 12,428 | +837 | 15729 | 77635 | 4023 | 1750 | 206 |
| Spain | 94,417 | +6,461 | 8,269 | +553 | 19259 | 66889 | 5607 | 2019 | 177 |
| China | 81,518 | +79 | 3,305 | +5 | 76052 | 2161 | 528 | 57 | 2 |
| Germany | 68,180 | +1,295 | 682 | +37 | 15824 | 51674 | 1979 | 814 | 8 |
| France | 52,128 | +7,578 | 3,523 | +499 | 9444 | 39161 | 5565 | 799 | 54 |
| Iran | 44,605 | +3,110 | 2,898 | +141 | 12656 | 27051 | 3703 | 531 | 35 |
| UK | 25,150 | +3,009 | 1,789 | +381 | 135 | 23226 | 163 | 370 | 26 |

Figure 28 - Snapshot of spread of COVID-19 on 31st March 2020

cold. Other members of the family, such as Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS), have more serious effects. The current virus is known as 2019-nCOV (2019-new COrona Virus) but it is more common to use the term COVID-19 (COrona VIrus Disease 2019) for both the virus and the disease it causes. COVID-19 is strongly related to SARS, is believed to have originated in bats (as do many other diseases) and to have infected other animals *en route* to infecting humans, although the exact transmission route is unknown.

The first known infections were detected in Wuhan, Hubei Province, China at the end of 2019. The 'timeline' for the virus is complex and still evolving; a brief summary is given below. At the time of writing, some three months after the initial cases were reported in Wuhan, most of the countries in the world have been affected and China is no longer the 'hot spot' for the infection as shown in Figure 28 drawn from [173], with Italy, Spain and the US having a higher death toll than China. Given the relative populations (60 million in Italy against 1.3 billion in China), the level of impact in terms of fatalities per million of population is vastly greater, as Figure 28, which gives a snapshot of the situation on 31 March 2020, shows. This is now truly a global phenomenon whose epicentre has moved around the globe – and it is far from over.

The spread of the virus has been astonishing. One of the key reasons for this is that the virus has a long incubation time and people infected by the virus can be asymptomatic (show no symptoms) but contagious, unlike with SARS. Therefore it can take several weeks before patients who contract the virus show symptoms, if indeed they do. This makes controlling the spread of the disease very difficult. Generally, the approach to controlling the spread has been to take measures to reduce human contact to slow the transference of the disease between humans. This includes restrictions on travel, 'social distancing' measures, such as keeping people apart, and recommending enhanced hygiene, such as more thorough washing of hands. The measures put in place have varied from 'recommendations' through to more stringent measures, including closing businesses, schools and universities, and limiting public gatherings. Many countries have imposed restrictions and slowly 'ramped them up' eventually reaching so-called 'lockdown' where people have to stay at home, only going out for essential reasons, such as to buy food or exercise. For example, Wuhan was in lockdown for some months (although some restrictions are being lifted at the time of writing) and some countries have closed borders to all but their own citizens. At the time of writing it was estimated that about 25% of the world's population was in lockdown.

The impact on the global healthcare system, or system of systems, has been enormous – but the true impact is much wider. There are widespread dependencies that have been exposed by the pandemic, for example, in sport, the Olympics being postponed to 2021, in entertainment, theatres and concert venues have been closed, and so on. Any one of these impacts is significant for the communities affected, but it is the cumulative impact that is more dramatic, and again shows the interdependence of apparently separate systems.

The impact on the economy has been enormous. Many businesses have seen a dramatic reduction in income and several airlines have already gone out of business. The lockdowns have meant that many businesses – for example shops selling non-essential goods – have closed. Some restaurants have moved to a delivery model

in response but, for many, there is no 'plan B'; for example, many musicians are self-employed and those who make their living through performance and teaching have seen their income reduced to near zero. Some governments have implemented economic stimulus packages and support for both companies and individual workers, but it is likely that some businesses that shut down 'temporarily' will never re-open. However, some businesses are actually benefitting, including delivery services and companies providing online conferencing facilities. There is both a short-term and a long-term impact – and the long-term may see very different business models than exist today, for example online conferencing reducing the need for business travel causing a long-term impact on transportation businesses. In the terms of complex systems, the COVID-19 pandemic may be seen as a tipping point or mode transition that will permanently change those industries.

Equally, there is a major impact on society. In many countries, individuals who are seen as vulnerable – usually the elderly and those with underlying health conditions that make them more susceptible to the virus – have been asked to self isolate. Those travelling have often had to go into quarantine, typically for 14 days. Many who can do so are now working from home (WFH) – and can perhaps be considered lucky by comparison with those whose livelihoods are now at risk. However, there is an impact on wellbeing of those in isolation or WFH, in terms of limited ability to take exercise and psychological effects. This is affecting both those who live alone and families who are both WFH and looking after children who can no longer attend school. Further, there is an impact on school children who are unable to take national examinations and are uncertain about their ability to progress to university. Challenging though these effects are, we are perhaps fortunate that this is happening now, rather than at the time of SARS (2002 to 2003). We have the internet and almost ubiquitous connectivity. Mobile phones can be used for video calls with friends and family, to order food deliveries, to contact emergency services if necessary and to support communications for those WFH.

From a complex systems perspective we are seeing emergent properties of what would hitherto have been thought of as independent (systems of) systems, which are now seen as being interdependent. While some of these emergent properties pose risk to health or safety, not just economic risks, some are positive, for example the reduction in the production of greenhouse gases and the improvement in air quality.

### *Understanding why it happened*

The pandemic has yet to run its course and there may be a recurrence of the disease as restrictions on movement are lifted or next winter when conditions are more favourable for spread of the virus. Thus, it is too early to give accurate reasons. What is needed is a retrospective analysis once the outcome is much clearer and this is one of our domain-specific recommendations. The aim would be to understand the management of the pandemic from the point of view of safer complex systems, in particular the impact on apparently independent systems. The detailed scoping of such a study cannot be determined now; however, it is possible, even at this stage, to identify some of the factors that need to be considered in this analysis:

- **Containment strategies** – Which strategies, implemented in different countries, for example the rate and the stringency of imposing lockdown and the effort put into tracking and testing the contacts of those known to have COVID-19, are most effective in terms of containing the spread of the infection and minimising fatalities?
- **Culture** – The extent to which citizens obey the guidance and/or instructions from their governments regarding social distancing and other measures to control the pandemic, and how this varies with other factors including religious persuasion, age (there is some evidence that young adults have been less willing to adhere to guidance), and income group.
- **Living conditions** – To what extent does the spread of the virus vary with living conditions, noting that it is common among poorer communities for multiple generations to live in close proximity, even in a single room, which makes social distancing impossible?
- **Economic conditions** – In some social groupings work is only available on a daily basis and the wage earners need to queue to get work (the alternative being to have no income and not be able to buy food) and this again makes social distancing difficult.
- **Healthcare systems** – What is the impact of the form of healthcare system? Noting that in some countries healthcare is largely provided by the state, in others it is provided privately (usually paid for by insurance) and in others, such as the UK and China, it is a hybrid.
- **Clinical resources** – The availability of clinical resources, or the rate at which new resources can be made available, including the ability to isolate those who have been confirmed with the disease, impacts both the quality of treatment and the spread of the disease.
- **Learning from experience** – To

what extent is experience of treating similar problems in the past, such as SARS, used to inform strategies for dealing with the virus? The Chinese authorities have produced guidance (in English as well as Chinese) based on their experience [174] with the intent of helping others learn from their experience.

- **Predictive models** – Deaths and other indications, see Figure 28, are lagging indicators and predictions are necessary to inform national policies and strategies for treating COVID-19. However, the accuracy of the models is crucial and there are some questions about the underlying assumptions of these models and the quality of their implementation, including the highly influential model developed by Imperial College London in the UK [175]. The quality of the models is very important given the far-reaching implications of the policies that are based on their predictions.

All of these factors are apparent now in the data, for example as shown in Figure 28, or in news stories from around the world, but hindsight will enable them to be evaluated more accurately and more dispassionately.

*Brief timeline of COVID-19*

China alerted the WHO to several unusual cases, initially thought to be a type of pneumonia, in Wuhan on 31 December 2019. At the time of writing, some three months later, COVID-19 has spread to be a worldwide pandemic, with far greater reach and impact than SARS or MERS. The following is a brief summary of the 'timeline' of COVID-19 over these three months. There are many more detailed descriptions of events; the one produced by Al Jazeera [176] is perhaps one of the most accessible.

Early analysis showed that the outbreak was not a recurrence of SARS and, as several of those initially infected had worked in the Huanan Seafood Wholesale Market, this market was judged to be the source of the infection and was shut down. The first death from the virus occurred on 11 January, and the first case outside China was reported on 13 January, in Thailand. By 20 January the virus had spread to at least 10 countries, and some, such as the US, had started to take actions such as screening people arriving from Wuhan. On 23 January air and rail departures from Wuhan were suspended and progressive restrictions were introduced across Hubei province, the first instance of 'lockdown' that has since become commonplace across the globe. By 25 January, some 56 million people in Hubei were under lockdown, and major restrictions were introduced including shutting down public transport and closing entertainment venues.

On 2 February the first death outside China was reported (albeit of someone from Wuhan). On 7 February Li Wenliang, a doctor who was a 'whistleblower' in Wuhan and later contracted the virus, died, showing the risks being undertaken by frontline medical staff. On 9 February the death toll in China exceeded that of SARS, with 811 deaths recorded and 37,198 infections. The virus continued to spread, with the first case in Africa (in Egypt) and the first death in Europe occurring (in France) on 14 February. On 19 February Iran reported two deaths, and on 20 February South Korea reported its first death. In late February new cases in China started to 'plateau' while the number of other countries reporting infections continued to grow.

Early March saw a rise in cases in the Middle East and Africa, including Jordan, Qatar, Saudi Arabia and Tunisia. On 7 March the worldwide total of cases exceeded 100,000 with the majority still in China. Several countries introduced lockdown, for example Saudi Arabia and Italy, including closing schools and universities. The WHO declared COVID-19 a pandemic on 11 March and, at the same time, the disease reached a number of countries in South America. On 18 March Italy recorded 475 new deaths, the highest death toll on any one day in any country (at that time), while China reported no new cases. By the end of March the total number of cases worldwide exceeded 800,000, with Italy, Spain and the US becoming the centres of the pandemic. It was estimated that about a quarter of the world's population was in lockdown, but in Wuhan restrictions were beginning to be lifted.

### C.3.3 Coronavirus and PPE

The COVID-19 pandemic has many different facets and it is not possible to address all of them in this report. However, it was decided that it was essential to give a concrete example of one non-epidemiological facet of the pandemic in the report. Personal protective equipment (PPE) is important for the safety of frontline clinical staff treating patients with COVID-19. This is a supply chain issue with a significant safety impact so it was chosen as a case study in the main body of the report, and is included as an illustration of the framework (see section 3).

# C.4 Case studies from the supply network domain

### C.4.1 Contamination of the food supply network by E. Coli.

In this case study we will draw some interesting comparisons between two examples of apparent contamination of food supply networks with E. Coli bacteria. The first case was in Germany in 2011 and the second potential case was in the UK in 2016. We will briefly introduce the two cases and then draw some comparisons.

*What happened?*

The German case was at the time the second largest incident on record with 4,321 cases, and the largest number of deaths at 50. The majority of the cases were in Germany, however there were cases reported in Austria, Czech Republic, Denmark, France, Greece, Luxembourg, Netherlands, Norway, Poland, Spain, Sweden, and the UK.

Initially there was a significant amount of confusion over the source of the contamination, highlighting uncertainty as a source of complexity in supply networks. Early reports linked the outbreak with Fenugreek seeds imported from Egypt by a German distributor in November 2009. However, it was pointed out by Egyptian officials that the contamination was unlikely to have survived for two years on the dried seeds. Hamburg also blamed cucumbers sourced from Spanish growers, which resulted in €210 million being paid out to farmers whose produce had to be thrown away [177].

The actual source of the contamination turned out to be beansprouts produced by a German organic farm in Lower Saxony. The beansprouts had passed the E. Coli test but still infected people, indicating a contributing failure in the regulation and test regime [178].

The contamination incident in the UK had some parallels but was on a different scale and ultimately evolved in a different way. Approximately 161 people fell ill from consuming salad in cafes and restaurants, two of whom later died. As such, the incident was of sufficient scale to be declared and managed as a national outbreak and reported to the WHO [179]. Analysis provided evidence that the consumption of mixed salad leaves, particularly from catering establishments, was associated with infection. However, sampling and microbiological testing of salad products at suspected sites of infection proved negative.

There was an additional suspected link with people who had recently returned from travel to the Mediterranean; however, this was ultimately proved to be inconclusive [180]. As a precaution, authorities halted the distribution of salad leaves imported from Italy, and Red Batavia from Italy was thought to be the most likely vehicle for the outbreak. No other European country reported a related outbreak, which could be an indication that the contamination was within the UK supply network or was introduced there. Public Health England eventually declared the outbreak to be over, and identified mixed salad leaves as the likely cause [181]. The eventual outcome was that Public Health England declared that no definitive source of the contamination was found [181].

*Why did it happen?*

Assuming that there was an incident at all (see discussion below) in the UK, then both outbreaks have a similar cause. A contaminate in the form of the bacteria E. Coli was able to 'leak' into the supply network and proved difficult to detect. In the case of Germany, the normal testing process was not able to detect the contaminate in the beansprouts that were eventually determined to be the source [178]. This very likely contributed to the spread of the outbreak and the initial uncertainty over the source of the contamination.

In the UK it was not possible to definitively trace the source, as the effects of the contamination were only clearly visible at the consumer level [181]. This lack of a definitive source does open up the possibility that there was in fact no single outbreak. Instead it is at least theoretically possible that what was observed was several isolated incidents that occurred over a similar time period. There could therefore have been no incident to stop.

The investigation of contamination in food supply networks uses a few different techniques. Epidemiological methods are used to try to determine if there is a pattern to the affected individuals, such as visiting the same restaurants or buying the same products. Microbiological techniques are used to determine the nature of the contamination, and what product(s) it is present in. Food safety and trace-back techniques are also used to try to follow the contamination back through the supply chain to its source [182]. It is hoped that the combination of these methodologies will allow the source to be traced through the system back to the source.

In reality this can be very difficult, as there is uncertainty around the information that can be gathered at each step. The uncertainty can be generated by limitations in testing, lack of responsibility for monitoring of the supply networks, or incomplete data; or, as in this case, a combination of those factors. This uncertainty hinders the ability to trace a source and also contributes to the incorrect, or complete failure of, attribution of the source, as was seen in both incidents. It also hampers our ability to judge if what is being observed is a perhaps convergent causality in a contaminated complex network, or just a coincidence [183].

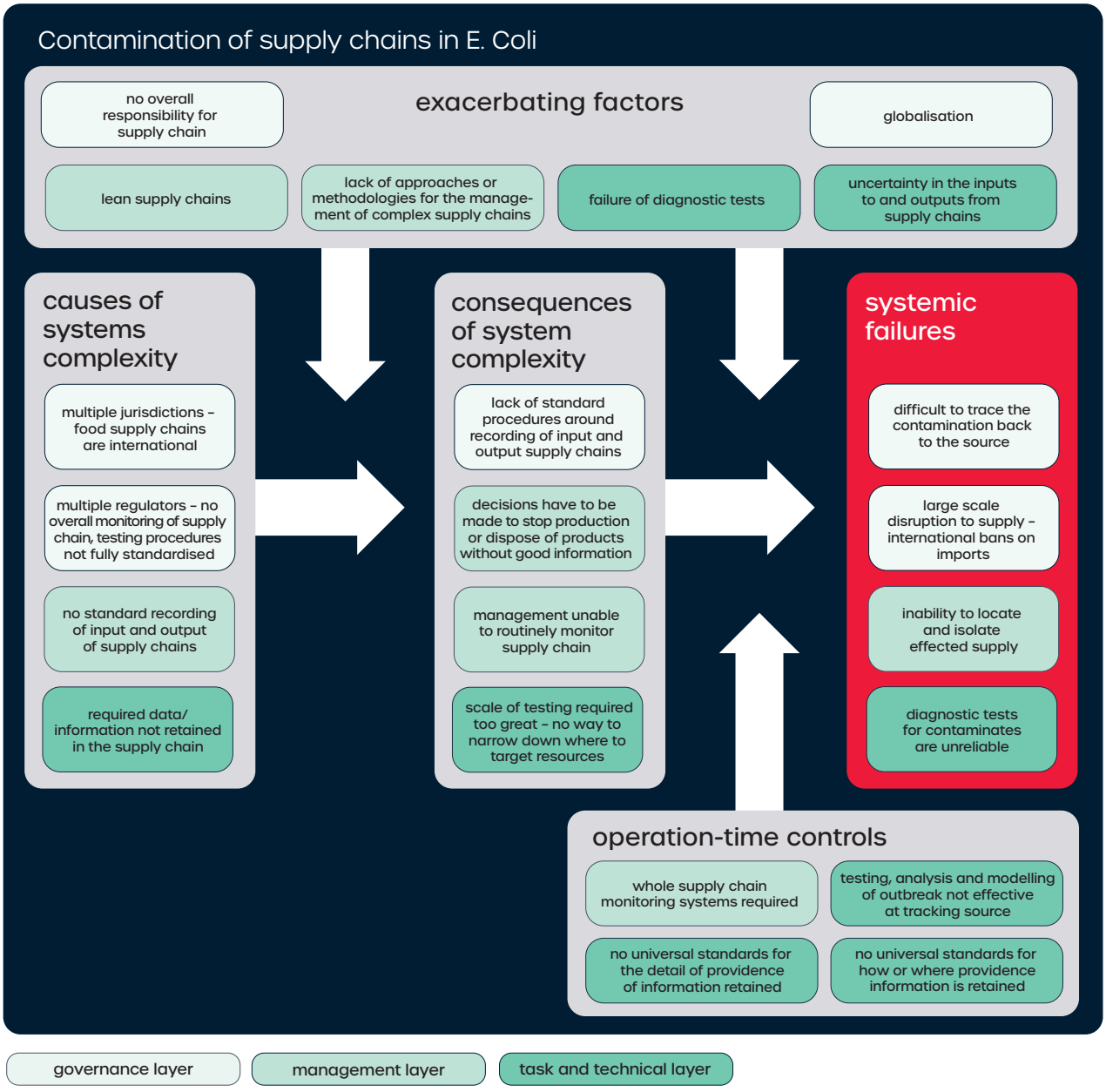



Figure 29: Analysis of factors in contamination of supply chains by E. Coli

### *The lesson for future complex systems operation*

Both incidents highlight a need for accurate information about the route of materials through supply networks; without this, the task of tracing contamination is all but impossible. The incidents also raise questions about where and how this information should be stored. As the complexity of supply networks has increased to the point where they behave more like complex networks, the complexity of monitoring all the inputs at any one point in the chain becomes increasingly difficult but also increasingly important. A summary of the complexities of the case are presented in Figure 29.

The incident in the UK, which is surrounded with uncertainty, highlights some of the difficulties in governing increasingly complex supply networks (raising questions about how this uncertainty could be controlled, and highlighting a need for better understanding of decision-making and human behaviour in complex dynamic environments [184]). This incident was only really evident at the consumer level and attempts to trace it through the system (across or through the network) to determine if the cases shared a common source failed. This failure should be viewed as a failure of governance, as in this instance we are not able to determine if there actually was a contamination, let alone trace that contamination to its source.

The incidents also raise questions about even if a source of

contamination is traced, are we able to determine how robust that process was and how reliable the analysis and its findings are? In the case of the German incident, the data on the outbreak was analysed until the analysis provided an explanation to what had been observed, and presumably no further. This does at least leave room for the possibility that the declared cause was not the true cause, as a new observation (piece of data) could invalidate it (which is the case for any hypothesis). In the UK the findings of the analysis of the incident could not be aligned with the observed pattern of the outbreak, and therefore no definitive cause was declared. What this suggests is that with supply networks that are complex systems it is particularly difficult to attribute causes of failure, because of the features of complex systems such as emergence and a lack of easily observed simple linear causality (see Appendix A.2). Therefore, there might be a tendency to not look beyond the first explanation that fits the observed pattern of failure. This problem is not unique to supply networks.

### C.4.2 PFAS forever chemicals

PFAS, the abbreviated form of perand polyfluoroalkyl, and perfluorinated alkylated substances, are synthetic organofluorine chemicals that have been used in a wide variety of products since the 1940s, and have a variety of chemical properties [185]. Their properties, which include lowering the surface tension of water, have led to their use in a range of consumer products, including cosmetics, food containers, polishes and other household products. They are also used in a range of industrial settings, including firefighting foams, oil production, mining, pesticides and clothing manufacture [185]. Their use came under increased scrutiny by the public and by scientific and regulatory communities in the 1990s, with significant resources being invested into research into understanding the effects of exposure to, and potentially controlling use of, long-chain PFAS in different countries [185]. This resulted in some forms of them being listed under the Stockholm Convention on Persistent Organic Chemicals and others being evaluated for listing [186].

The reason that in the latter half of the 20th century interest grew in PFAS chemicals was that it became clear that they were among a relatively small group of chemicals that persist in the environment, potentially forever. This is of concern as chemicals that persist, and do not breakdown, will naturally accumulate in the environment. Furthermore, some persistent chemicals, including PFAS, also bioaccumulate (remain and increase in concentration in biological tissues in, for example, seafood [187]) as organisms are unable to excrete them or break them down. Bio-accumulation of chemicals is obviously concerning, as it raises the question of what the impact on the organism is over the long and short term.

One of the ways which PFAS chemicals get into the environment, and are therefore able to 'leak' into other systems, is through their use in supply networks and the production of products.

#### *Firefighting foam*

One significant source of environmental contamination by PFAS chemicals is from the use and production of some types of firefighting foam, in particular aqueous film-forming foam (AFFF). AFFF are used in fire extinguishers; however, a much larger source is their use in the aviation and defence sectors for aviation rescue firefighting, particularly as part of a training regime to prepare for aircraft accidents.

Often sold under the brand name 3M Litewater, AFFF chemicals were used extensively throughout the 1980s and 1990s in aviation rescue firefighting, both in training and incidents, as well as for the extinguishing fires involving highly flammable liquids. In some countries it was a regulatory requirement, particularly in the case of aviation rescue firefighting, that the AFFF chemicals were used during training. This resulted in large amounts of the chemicals entering the environment [188], and there is evidence of bio-accumulation in the firefighters who were exposed to the chemicals [189]. AFFF chemicals have been found in drinking water, fish, soils, and even in Barents Sea Polar Bears [190, 188].

There have also be been more specific examples of environmental and health concerns associated with PFAS chemicals and their use in airports. In early 2017 Qantas advised the Queensland Government in Australia that firefighting foam had been spilling near Brisbane airport. Despite stating that there was no consistent evidence linking PFAS chemicals to health dangers, the health authorities warned recreational fishers to avoid the area, and suggested that residents avoid eating seafood from nearby areas.

#### *Why did it happen?*

The cause of what can broadly be called environmental contamination by PFAS chemicals used in the firefighting supply network could perhaps be attributed to lack of regulatory foresight and then action, combined with insufficient scientific understanding of the properties of the chemicals (particularly early in their use). Initially, the potential for the environmental damage, and particularly bio-accumulation, was not and perhaps could not be known. These were a new form of artificial chemical, not seen before in nature. It is therefore difficult to understand how they

will behave in nature over the long term - they are an intervention in a very complex system. However, one could argue that introducing anything that is unknown to the environment, into the environment, comes with significant risk. This story then becomes one of when does the scientific evidence become sufficiently strong that the regulators should act to ban or limit the use of PFAS chemicals, or something similar? It also raises wider questions around how society should approach the use of novel technologies.

An additional contributing factor is that there are two significant competing objectives that perhaps delayed (and complicated) action to stop or limit the use of AFFF for aviation rescue firefighting: the ability of rescuers to prepare effectively for fighting aircraft accident fires being one, and the potential for environmental damage.

In 2003 Australia's federal chemical regulator issued a warning about the possible environmental and health dangers of PFAS, following advice from the US in 2003. At the same time, due to concerns, the sale of firefighting foams containing a particular PFAS, perfluorooctane sulphonic acid (PFOS), ceased in 2003; but many AFFFs being sold still contained another PFAS, perfluorooctanoic acid (PFOA) and its precursors [191].

Attempts to reconcile the competing objects of accident preparedness and AFFF contamination impact can be seen in new regulations, such as those set out in a 2015 policy overview conducted by the Queensland Government, which provides updated foam management and use standards [192]. Those updated policies are not enforced until 2019, some 15 years after the initial concerns led to the removal of PFOS from AFFF.

This case highlights a temporal aspect of the safety of complex supply networks. The problems of PFAS chemicals became apparent over time as a better understanding of their behaviour in the environment was developed (and continues to be developed). This presents a challenge: who is liable for any damage to the environment? Does it lie with the company or does it lie with the regulators? More often than not, it is taxpayers who have to pay for any clean-up. Regulators do provide some of the context, or the environment, in which companies operate, and shape some of their operations; this also gives them responsibility to limit the use of technologies that are found to be harmful, and perhaps they also have a role in limiting the use of technologies about which little is known.

The case study highlights a significant challenge in tackling safety issues with the complex natural environment when two imperatives are competing: the need for firefighters to tackle blazes effectively and safely, and the need to avoid long-term environmental damage.

# C.5 Case studies from the railway domain

### C.5.1 Hatfield rail crash

Four people were killed and more than 70 people injured, four seriously, when a Leeds-bound GNER train carrying 170 passengers derailed south of Hatfield station[193].

The immediate cause of the derailment was the fracture and subsequent fragmentation of the rail near to Hatfield. The rail failure was due to the presence of multiple and pre-existing fatigue cracks in the rail. The underlying causes identified by the HSE investigation were that the maintenance contractor at the time, Balfour Beatty Rail Maintenance Ltd (BBRML) failed to effectively manage the inspection and maintenance of the rail at the site of the accident. The investigation also found that Railtrack plc, the infrastructure controller at the time, failed to effectively manage the work of BBRML. In the terms of the framework causes included "no single owner" and "collaborating hierarchically managed systems" with systematic failures including "accountability mismatch" as the responsibility for ensuring safety was distributed among several organisations.

### C.5.2 Crash on the MTR Tseun Wan line in Hong Kong

A joint venture between Alstom and Thales was installing a new signalling system on the MTR Tseun Wan line in Hong Kong in 2019. On 18 March while testing the new system in non-traffic hours, two trains collided causing damage to carriages and a derailment [194]. The train drivers were taken to hospital for checks but were released on the same day.

The accident occurred while testing the 'fail-over' of the signalling system, specifically a Zone Controller (ZC). The systems traditionally used for ZC are duplex and a failure in one channel will cause a switch over to the other. The client required a triplex design to be used and a warm-standby tertiary ZC was developed. The standard duplex design is mature, but new software was developed for the nonstandard tertiary systems and this failed to re-create correctly data for a section of track near Central Station, which prevented the automatic train protection (ATP) system from operating. In the terms of the study framework, the causes include "path dependency" and "system evolution".

### C.5.3 Kings Cross Station fire

At 7.45pm on 18 November 1987, a fire started at King's Cross St Pancras station, a major interchange on the London Underground [195]. It was the worst fire in the history of the London Underground, with 31 people dying and many more seriously injured. The fire began in a mixture of grease and debris that had accumulated on the running tracks of the Piccadilly line escalator number four during its entire operating life. It was later found that the escalator had never been completely cleaned since being installed in 1939. Ignition was attributed to a discarded match of a smoker, which fell between the tread and the skirting board of the escalator. Even though smoking is banned in the Underground, passengers were known to light up while riding the escalators to the surface. In task and technical terms, the proximate cause of the fire was a "human–system interaction" but the underlying causes lie in the managerial layer and the systemic failure is an "unanticipated risk".

# C.6 Case studies from the oil, gas and chemical process industries

### C.6.1 Piper Alpha

Late in the evening of 6 July 1988, a series of explosions ripped through the Piper Alpha platform in the North Sea. Engulfed in fire, over the next few hours most of the oil rig topside modules collapsed into the sea. 167 men died and many more were injured and traumatised. The world's biggest offshore oil disaster affected 10% of UK oil production and led to financial losses of an estimated $2 billion [196]. Many of the elements of the study framework can be seen in this disaster, not the least of which was 'competing objectives' where the pressure to continue production contributed to the seriousness of the fire. One of the consequences of the disaster was the introduction of a safety case regime and the use of safety cases has now spread into many other industries.

### C..6.2Deepwater Horizon

On 20 April 2010 the mobile offshore drilling unit (MODU) Deepwater Horizon was completing drilling operations at the Macondo well oil exploration project in the Gulf of Mexico on the US Outer Continental shelf, preparing to temporarily abandon the well. During these operations, there was a loss of well control that resulted in a release of liquid and gaseous hydrocarbons, which culminated in explosions, fire, the loss of 11 lives, the eventual sinking and total loss of the Deepwater Horizon, and the continuous release of hydrocarbons into the Gulf of Mexico [197]. The flow was stopped on 15 July 2010 and the well was declared sealed on 19 September 2010.

The Republic of the Marshall Islands Maritime Administrator concluded that the proximate cause of the casualty was a loss of well control resulting from:

- Deviation from standards of well control engineering.
- Deviation from the well abandonment plans submitted to and approved by the Minerals Management Service (MMS).
- Failure to react to multiple indications that a well control event was in progress.

Non-causal factor conclusions included:

Better communication and coordination between the flag state and the coastal state regarding inspections and surveys could help to ensure that both the flag and coastal states are aware of conditions or requirements that could affect the safety of MODUs and their personnel.

The unit withstood the forces of the explosions and resulting fire, providing a sufficiently stable and protected platform to facilitate the evacuation of 115 of the 126 persons on board.

The electrical power failed at the time of the first explosion or immediately thereafter. The failure of the primary power source added to the confusion during evacuation and complicated the evacuation of the unit.

The total loss of electrical power compromised the functioning of the fire suppression systems; however, any attempts at suppression would have been futile given the intensity and magnitude of the fire and the uncontrolled fuel supply. It is unlikely that any ship-borne system would have been effective at extinguishing the fire onboard the Deepwater Horizon.

The Emergency Disconnect System did not function as intended and the unit was unable to disconnect. Without any ability to stop or reduce the flow of hydrocarbons, and without power for vital systems, the crew was forced to evacuate the unit.

There were instances of confusion regarding decision-making authority during the casualty. While such instances highlight the fact that the integration of drilling and marine operations presents challenges for maintaining a clear command hierarchy, especially in emergency situations, there is no indication that any confusion as to the chain of command was a causal factor in the casualty.

Ideally, the evacuation of a unit occurs in phases. However, the speed at which the casualty progressed provided limited time for reaction, control, mitigation efforts and response. That 115 individuals were able to safely evacuate the Deepwater Horizon is due in part to the robustness of the underlying regulatory system, including requirements for redundancy of lifesaving equipment, routine fire and emergency drills, and safety orientations for all visitors to the unit.

The proximity of the Damon B. Bankston (supply ship) and the timely and effective response of its crew substantially contributed to the successful evacuation of the Deepwater Horizon.

# C.7 Case studies from the military domain

### C.7.1 Black Hawk friendly fire incident

On 14 April 1994, two US Air Force F-15 operating under the control of a USAF airborne warning and control system (AWACS) aircraft, misidentified two US Army UH-60 Black Hawk helicopters as Iraqi Mil Mi-24 'Hind' helicopters. The F-15 pilots fired on and destroyed both helicopters, killing all 26 military and civilians aboard, including personnel from the US, the UK, France, Turkey, and the Kurdish community [198].

### C.7.2 Nimrod XV230

The UK Royal Air Force (RAF) Nimrod XV230 was on a mission over Helmand Province in Southern Afghanistan on 2 September 2006 in a ground-support role when it suffered a catastrophic mid-air fire, leading to the total loss of the aircraft and the death of all those on board. The technical causes of the accident included modifications to legacy systems ('path dependency') without a proper analysis of the safety consequences. Sir Charles Haddon-Cave's review of the accident [199] was far-reaching and highlighted, among many other things, the management failings that contributed to the accident. In the report and in subsequent talks Sir Charles particularly emphasised the importance of "responsibility" and clarity in the "duty holding" construct. He also referred to "confirmation bias" in safety work in assuming that the aircraft was safe as it had been flying for many years, so the aim in producing the safety case was simply to show that the aircraft could continue to operate.

Sir Charles also discussed many issues that fit into the 'operation-time controls' in this report's model, including discussing safety culture, a crosscutting issue, and learning from experience. He also talks about "comfort in complexity", such as "an organisational structure which is of Byzantine complexity can look impressive in a coloured organigram or PowerPoint but is likely to indicate diffuse responsibility, attenuated lines of accountability and confusion... as to who does what". The report – long though it is – is worth reading carefully as it holds salutary lessons in what can go wrong, and some strategies for resolving the problems. He stresses the need for simplicity – which we also see as one of the reasons why systems have been so safe in the past.

Sir Charles Haddon-Cave's review of the accident talks about "comfort in complexity", which can seem impressive but is likely to indicate diffuse responsibility, attenuated lines of accountability and confusion.

# C.8 Case studies regarding responses to natural disasters

### C.8.1 Australian bushfire preparedness

One of the expected consequences of climate change is an increase in the frequency and intensity of weather extremes such as heatwaves, droughts, and large-scale bushfires. The possible escalation in the frequency and magnitude of resulting impacts has led to arguments that future strategies for emergency management should be based on achieving organisational and community resilience. However, relatively little is known about the limits of conventional emergency management approaches and factors leading to resilience. Drawing on the 2009 bushfires in the state of Victoria, Australia, as an analogue for a 'more-severe-than-expected' event likely under a future changed climate, this article [200] analyses the limits to emergency management approaches under unfamiliar conditions. The assessment focuses on three organisations involved in the Victorian bushfires emergency response. Results show how events that occur with unprecedented severity are well beyond the routine emergency management capacities of emergency organisations. The paper discusses how the long-term promotion of organisational and societal resilience could be achieved and outlines implications for research and practice.

### C.8.2 Hurricane Katrina preparedness and response

When Hurricane Katrina made landfall near the Louisiana-Mississippi border on the morning of 29 August 2005, it set in motion a series of events that exposed vast numbers of Americans to extraordinary suffering [201]. With the breaching of the levees, the city of New Orleans flooded, requiring the emergency evacuation of tens of thousands of residents who had not evacuated before the storm. Lifted off roofs by helicopters or carried to safety in boats, they were taken to the Superdome, the Convention Center, a piece of high ground known as the Cloverleaf, and other dry spots around the city. At these locations, they were subjected to unbearable conditions: limited light, air, and sewage facilities in the Superdome, the heat of the sun, in many cases limited food and water, and fear for their personal safety and survival – and the survival of their city.

The possible escalation in the frequency and magnitude of natural disasters and the impact of environmental change has led to arguments that future strategies should be based on achieving organisational and community resilience.

# C.9 Case studies from the built environment

### C.9.1 Lancaster power outages

In December 2015, life for more than 100,000 people in Lancaster reverted to a pre-electronics era. A flood at an electricity substation resulted in a blackout over the entire city that lasted for more than 24 hours. Suddenly people realised that, without electricity, there is no internet, no mobile phones, no contactless payment, no lifts and no petrol pumps. Although these dependencies were not difficult to see, few had thought through the implications of losing so many aspects of modern life at once [89]. This example illustrates the importance of understanding the interdependencies between systems and of making systems resilient to failure.

### C.9.2 Grenfell Tower

In the early hours of 14 June 2017, a fire spread through Grenfell Tower, in London. Seventy-one people died, many homes were destroyed and countless lives were affected. The fire appeared to be accelerated by the building's exterior cladding system, leading to a national programme of extensive testing of the cladding on other high-rise buildings. This revealed widespread use of aluminium-composite materials that did not meet the limited combustibility requirements of building regulations guidance, and raised concerns for the safety of other buildings. Further concerns soon came to light about the adequacy of the structural design of cladding systems when materials fell from a building in Glasgow. A subsequent series of fire and rescue service audits of tower blocks led to the temporary evacuation in London of the Chalcots Estate, Camden, and resulted in the discovery of structural safety issues with four buildings at the Ledbury Estate, Southwark.

The key issues underpinning the system failure included [138]:

- **Ignorance** – Regulations and guidance are not always read by those who need to, and when they do the guidance is misunderstood and misinterpreted.
- **Indifference** – The primary motivation is to do things as quickly and cheaply as possible rather than to deliver quality homes that are safe for people to live in. When concerns are raised, by others involved in building work or by residents, they are often ignored. Some of those undertaking building work fail to prioritise safety, using the ambiguity of regulations and guidance to game the system. This is an example of 'conflicting objectives' in the terminology of the framework.
- **Lack of clarity on roles and responsibilities** – There is ambiguity over where responsibility lies, exacerbated by a level of fragmentation within the industry, and precluding robust ownership of accountability.
- **Inadequate regulatory oversight and enforcement tools** – The size or complexity of a project does not seem to inform the way in which it is overseen by the regulator. Where enforcement is necessary, it is often not pursued. Where it is pursued, the penalties are so small as to be an ineffective deterrent.

The report also makes wide-ranging recommendations for improvement. For example, the Health and Safety Executive (HSE) has now been charged with implementing the Building Safety Regulator (BSR) and is in the early stages of doing so.

# C.10 White goods

As mentioned in the stakeholder workshop, it became clear that 'white goods', for example domestic appliances such as fridges and washing machines, which though simple in themselves operate in a surprisingly complex environment. The aim here is not to highlight a particular accident or incident (although we note that the failure of a fridge-freezer was the trigger for the Grenfell Tower fire), but rather to illustrate the complex ecosystem surrounding white goods. Factors in the wider ecosystem include:

- **The business model affects willingness to register** – the details of owners are sold on, which makes people unwilling to register so goods are hard to track.
- **Consumer to consumer sales (including selling on recall items)** – this further exacerbates the difficulty of tracing goods and ensuring that safety recalls are implemented.
- **Lack of traceability in the supply chain** – it is difficult to identify where parts have come from, or if counterfeit parts have been supplied, and this can hinder rectification of safety problems.
- **Right to repair** – meaning that people other than the manufacturer can repair the products, so it becomes ever more difficult to know the 'build status' of the products.
- **Impact of net zero** – seeking to have a zero-carbon footprint means the products are produced and recycled in a very complex ecosystem.
- **Use of white goods at night** – this is 'encouraged' by net zero, meaning that products are unsupervised while operating or charging, giving rise to a greater fire risk (the risk of starting the fire may be no different but the likelihood of detecting it quickly is lower).
- **Circular economy** – this is related to net zero and is concerned with eliminating waste, so has an impact on the design and manufacture of the products.
- **IoT white goods** – use of IoT may help tracing, but it increases the technical complexity of the products, and may result in surprising behaviour for consumers.
- **Cybersecurity** – as products are internet-connected it is possible to 'hack' them and, for example, turn on cookers in the middle of the night.
- **Data and security/privacy** – especially where devices are internet-connected, it is possible for third parties to access the devices and learn about their owners' habits for commercial exploitation and potentially compromising fundamental rights of privacy. This has included manufacturers such as Samsung warning their customers that an audio stream collected for the purposes of voice-activated SmartTV commands could be passed onto third party companies [202].

These complexities leave us with the question of who is responsible for the product. This may be an area where legislative changes are needed to ensure that such white goods are managed responsibly. Although white goods are comparatively simple in the task and technical dimension, they are a lot richer in the governance and management dimensions.

D

# Sector-specific recommendations

This section of the report gives recommendations for aerospace, mobility, healthcare and supply sectors.

# D.1 Aerospace-specific recommendations

As discussed earlier in this report, the aviation industry has a mature approach to safety management. However, the future rapid changes in technology and services within the industry suggest that there will be a need to improve approaches to managing complex systems. Proactively working in the areas outlined below will allow the industry to prepare for these changes and maintain the successful track record of safety improvement.

The following recommendations are focused on framework elements that are believed to require the greatest focus for **Safer Complex Systems:**

- **A1**: Ensure global and national organisations have frameworks to learn lessons and adapt quickly following complex system failures and accidents. This should occur at all layers.
- **A2**: Prepare means to respond to rapid technology change at the governance layer to prevent regulation lag.
- **A3**: Ensure frameworks at the management and governance layers are protecting against accountability and moral responsibility gaps.
- **A4**: Ensure frameworks at all layers consider means to manage emergent behaviour in the operation of complex systems.
- **A5**: Ensure the governance layer applies SMS and safety case techniques in a manner similar to the management and task/ technical layer.
- **A6**: At the governance layer, consider means of managing cross-jurisdictional issues that may impact the effectiveness of safety regulation.

The aerospace sector has learnt from incidents and accidents over decades to drive down aircraft accident rates but changes in governance and management are needed to preserve this trend in the face of technology change and the growth in system complexity.

# D.2 Mobility specific recommendations

The automotive industry and ground-based transportation in general is currently in the midst of a radical period of transformation. Mobility services increase in line with trends away from personal ownership of combustion-engine-based personal vehicles, bringing about a more connected perspective on the role of mobility within society. This trend is most obvious in the popularity of car sharing services and in increasingly multi-modal forms of transport (including e-scooters, delivery drones and automated people movers). Technical innovations in the areas of personalisation, automation, connectivity and electrification make these changes possible. Artificial intelligence, and more specifically machine learning technologies are also seen as an important step towards achieving human-like (and arguably super-human) levels of performance in automated driving systems. Many of these innovations are specifically targeted towards reducing traffic accidents and emissions. However, new classes of risks are also introduced that can no longer be assessed at the vehicle-level alone and require an understanding of the systems-of-systems interaction within the entire mobility infrastructure and its context to demonstrably achieve the safety potential behind these innovations.

Automation is not the only driver of complexity in the mobility sector, for an example, see the smart motorways case study. However, automated driving systems, both as a function within a standard passenger vehicle (car) or as part of a wider system (such as a people mover or tram), provide a stark demonstration of the relationship between increasing system complexity and safety. In particular, the following observations can be made:

Technically perfect automated driving systems will not be feasible (at least not in the foreseeable future). Safe deployment will therefore depend on measures at the management and governance layers.

AI-enabled automated driving systems exhibit properties of complex systems. Understanding the causes and effects of complexity across system layers is therefore key to managing the safety of the overall system.

A systems-oriented approach that acknowledges complexity and includes coordinated measures across governance, management and task and technical layers is required. This will require closer collaboration between domains such as automotive manufactures and suppliers, communication providers and city infrastructure as well as a better understanding of dependencies across the three layers of the framework. These recommendations are therefore primarily targeted towards industry and regulatory organisations. For example, this could include providing input to the UK's Department for Transport's Center for Connected and Autonomous Vehicles (CAV) as well as direct engagement with industrial partners and international standardisation bodies. However, progress in this area must also be supported by research activities (see Section 7.2) as many fundamental challenges are yet to be solved.

- **M1: Definition of safe for automated driving and interconnected mobility services** – Consensus should be developed for safety targets for automated driving and new forms of ground transportation that rely on a high level of inter-connectivity to other services and infrastructure. This should consider both quantitative measures (for example based on accident statistics) as well as qualitative approaches (based on engineering practices and operation-time controls) for achieving acceptable levels of risk. Achieving consensus will require cross-disciplinary dialogue involving not only technical but also legal and ethics experts. This is required to reach a level of trust and acceptance of the systems, without which the safety benefits of increased automation will also not be realised. Wider engagement with the public in general is also required to consider the perspectives of those most impacted by risk, and also to gain an understanding of the expectations and assumptions made on the systems by the users.
- **M2: Informed, outcome-based regulation** – Due to the rapid technological changes driving the transformation of the mobility sector, it is not feasible to expect that traditional approaches to standards development will keep pace with the rate of change. Therefore it is proposed that outcome-based regulation that stipulates requirements on *what* to argue instead of *how* to argue safety is developed; it should take a systems-oriented view with additional focus on arguing the effectiveness of controls for reducing risk due to system complexity. Published standards and regulations should be supported by publicly available specifications that provide more specific guidance and document current industry consensus on topics such as assurance activities for machine learning in an automated driving context. These specifications can be developed in a more agile manner than full standards and can therefore be continuously updated to reflect state-of-the-art.
- **M3: Operation-time controls and continuous assurance** – Ensuring safety of current automotive systems currently places a strong focus on design-time controls and type approval. However, as the complexity and scope of the

systems increases, and with it the sensitivity to an ever evolving environment, it is unrealistic to believe that an adequate level of safety can be achieved before the system is deployed and can be maintained over the vehicle's lifetime. Statistical arguments based on miles driven between incidents during field-based tests become both unfeasible and ineffective due to the effort required to collect the data and the difficulty in ensuring sufficient coverage of edge cases and critical situations. The increase in use of simulation during the design and validation of the systems allows for a more targeted testing of critical and rare situations. However, such approaches require additional arguments regarding the accuracy and transferability of the results into the target domain. Manufacturers, operators and regulators must therefore agree on a set of operation-time measures for ensuring the safety of the systems that includes the measurement of critical observation points within the system (leading indicators of systemic failures) as well as whether assumptions made regarding the operational design domain and therefore the validation approach continue to hold. The assurance case for the system should be continuously evaluated and refined, based on experiences in the field and changing expectations on the system. This holds true for automated driving applications but also to connected traffic infrastructure in general.

- **M4: Holistic safety (and security) analysis and risk management methods** – Safety analysis methods within the automotive industry were previously focused on analysing the occurrence and propagation of faults at a technical component level. The industry must support the development and adoption of systematic risk analysis methods at a system (of systems) level that include the vehicle, the supporting infrastructure and its environment, taking into account the impact of complexity at the task and technical as well as management and operations level. The fault model supporting these analyses should be expanded from that currently considered within functional safety standards which focuses on random hardware failures and systematic design failures (such as software bugs). This will include broader categories of technical causes of systemic failures such as cyberphysical attacks, functional inefficiencies of components and gaps in understanding of the operational domain. However, causes at the management/operational level, such as deliberate or accidental misuse, inadequate monitoring of performance in the field, and conflicting objectives resulting from regulatory constraints and societal expectations should also be considered.

Furthermore, the scope of these analyses must not be restricted to individual components or functions within the vehicle. Analyses need to be applied that include the traffic infrastructure, connected services, other traffic participants and the environment. The objective of such analyses should be to increase the robustness and resilience of mobility systems to the effects of complexity and lead to a managed level of risk associated with systemic failures of the system. The resulting analysis strategy will require a combination of existing and novel methods best focused towards different properties of the system.

- **M5: Systems engineering approaches to traffic infrastructure** – Related to M4 above, a systems-level approach to co-designing traffic infrastructure in line with new modes of transport and levels of automation is required. In doing so, the strengths and weaknesses of new forms of transport and technology should be taken into account including their interactions with other road users including, for example, cyclists and pedestrians (see Appendix C.2.2). As was demonstrated in the Uber Tempe case study, subtle effects in the environment (ambiguous nature of the piece of ground separating the directions of the carriageway) can encourage behaviour that may not have been anticipated in the design or operation of the vehicle. Taking a broader view of the system under design will also allow for additional solutions to be found that can lead to more robust and resilient systems. As an example, the use of infrastructure to monitor traffic flow and signal stop lights may be more accurate, resilient to component failures and cost effective than any advanced perception systems that could be mounted on a vehicle. However, increasing interconnectivity within the system will also inevitably lead to additional emergent properties that cannot yet be anticipated. This level of infrastructure/vehicle co-design will require a high level of cross-industry collaboration and support at the governance level where appropriate standards and regulatory approaches are either currently disjointed or lacking altogether.
- **M6: Understand and quantify the limits of AI** – An essential prerequisite for applying AI techniques, and in particular machine learning for safety-critical tasks in automated driving, is an understanding of the performance limitations so that appropriate system-level measures can be applied to counteract any functional inefficiencies. The focus here

should be on domain-specific guidelines on training data collection and demonstrably effective verification techniques including wide-scale field-based studies to support theoretical analysis and the development of standards and publicly available specifications of best practice (see M2). Research in the field of verification and validation of safety-critical ML applications is immature and more work is needed to demonstrate how ML-specific performance measurements and analyses correlate to the overall system safety goals [203].

- **M7: Manage the complexity of automated driving in line with confidence in the safety arguments** – The capabilities required to safely deploy automated driving systems will need to be developed and confirmed over time, thereby limiting the speed at which the systems can be introduced into the market. This is due to several factors:

  - The need to develop competencies in system safety methodologies for open-context autonomous systems within the automotive industry, including a significantly strong foundation in basic systems engineering principles.
  - The need to resolve open research questions that are required for a convincing safety assurance case.
  - Technological development of the tool chains and infrastructure required for design, simulation and test of the systems.
  - The efficacy of the methods referenced within the safety assurance case must be confirmed for realistic examples (for example the ability of innovative testing techniques to demonstrate the robustness of machine-learning-based perception functions).
  - Pre-validated system components with known functional and performance properties must be developed for re-use that can be applied to successively more sophisticated functions without requiring a complete system re-validation.

  The industrialisation of the assurance approaches for large-scale series development and release of such systems will require major changes across the industry. An iterative approach to developing these capabilities and confirming their effectiveness is therefore recommended. It is therefore recommended that a carefully managed, monitored and regulated approach to deploying autonomous driving functions is taken. This will include restricting functionality and operating scope to such levels where the risk can be managed with existing methods and introducing a systematic approach to gathering data on the effectiveness of safety management and the emergence of as yet unknown interactions and risks in the traffic system. Such an iterative approach could include:

  - Increasing the level of autonomy – from hands-on lane-keeping assistant functions to hands-off traffic jam partial automation and increasing towards autonomous driving on specific sections of motorway. This would involve addressing the complex human factors issues related to handover of control between vehicle and driver.
  - Increasing the complexity of the operating domain, for example from restricted weather conditions or geo-fenced areas with strict controls (such as cargomovers in closed-off port environments) towards inner-city driving under all normal weather conditions and times of day.

# D.3 Healthcare-specific recommendations

As noted in the main body of the report, healthcare is a system under pressure, and this is all the more evident given the global impact of COVID-19. There are, of course, many immediate responses to COVID-19, such as seeking to develop a vaccine and accelerating the production of vital equipment such as ventilators. This report's aim here, however, is to extract recommendations that are relevant from a complex systems standpoint and not to be limited to issues that relate to COVID-19 – although they do illustrate some of the real challenges for the sector.

The following recommendations are aimed primarily at healthcare providers and regulators; technical support will be required for some of the recommendations, but responsibility should lie with service providers and the regulatory community.

- **HC1: COVID-19 retrospective** – At a suitable time – perhaps once it is clear if COVID-19 will recur as a pandemic or not in the winter of 2020–21 – conduct a thorough retrospective analysis of the management of COVID-19. This should focus on the complex systems issues, not epidemiology, comparing and contrasting different strategies in different countries, including addressing the factors identified in Appendix C.3.2. It should consider, among other things, risk identification, risk stratification, risk transfer (including between socio-economic and ethnic groups) with the aim of learning effective and equitable risk management and control strategies before the next such pandemic. It should also consider the positives from the pandemic, such as international collaboration like publication of the results of Chinese experience in English [174], fast publication of research results for the benefit of all, changes in emergency room flow, and the willingness of companies and individuals to work for the public good (see Section 3.3) – to see how such benefits might be sustained in the future.
- **HC2: Safety management** – investigate how to enable healthcare to benefit from both traditional safety engineering ('Safety I' – with its focus on learning from and avoiding errors) and more recent approaches to managing safety (broadly 'Safety II' – with its focus on learning from normal or exceptional performance), finding appropriate ways to combine and balance the approaches, drawing on existing work, for example [204, 68, 55]. This might, for example, involve blending Hollnagel's notion of variability with methods for deviation and fault/failure analysis such as HAZOP, FMEA and FTA (Safety-I methods). In doing this it will be necessary to consider human factors in terms of understanding and communicating risk, trust in technologies (especially if incorporating AI), and how to obtain effective feedback from operations to help reinforce the ways in which things go right, perhaps using ML to understand work as observed as opposed to work as imagined (or defined) [205]. Ultimately this might result in establishing 'Safety-III' embracing a broader understanding of human factors – how people behave and (mis) behave – and a socio-cultural perspective. To realise such improvements also needs changes to the training of all involved – hospital managers as well as clinicians.
- **HC3: Assurance of models, AI/ML and autonomy** – Develop appropriate standards and guidelines for the development and assurance of models, AI/ML and autonomy used in critical situations, in order to support effective regulation. This should cover models used for decisions affecting individuals, for example for sepsis treatment, and those that can affect whole populations, such as the spread of viruses. The approach should recognise different types of model, as illustrated below, with the assumption that there would be progressive (cumulative) requirements for safety assessment and assurance:
  - **Statistical models** – define controls over the data sources, data analysis, and software engineering standards to ensure the validity and trustworthiness of the models used in support of decision-making; ensure that the models used take into account human factors and fit into the clinical workflows, drawing on relevant guidance, for example [206].
  - **AI/ML** – ensure control over the AI/ML development, covering data preparation, model learning and model verification, drawing on existing work that shows how to integrate assurance into the AI/ML lifecycle, for example [207]; carry out a hazard and risk assessment (see autonomy below).
  - **Autonomy** – where systems can act autonomously (without direct human control), conduct a complete hazard and risk assessment, producing and implementing derived safety requirements for the autonomous elements (whether using AI/ML or not), summarising the results through a safety case, drawing on both Safety-I and Safety-II precepts.

  A critical element of all this is model validation, recognising that all models are imperfect but that it is important to understand the gaps between intent and what is actually specified [1] in order to assess utility and validity. It may be appropriate to build on work by the US FDA on AI and ML in software as a medical device (SaMD) [208] and the regulatory sandbox undertaken

by the CQC and MHRA [103]. There also needs to be an incremental approach to introducing such technology, both to control risk and to enable clinicians to build trust in the technology. Note that any such approach would also need to consider operational monitoring of deployed systems and use of feedback from operations to improve the safety of the technology [115].

- **HC4: Open information management** – Healthcare is data rich and, in principle, it is possible to use the data to learn from experience and to improve safety. Often, access to data is slow and difficult due to delays in publication, patient confidentiality issues, differences in record-keeping and data formats used in different countries and by different healthcare system providers. This is a hindrance to carrying out research that could improve patient safety. However, COVID-19 has shown some very encouraging trends in fast publication and information sharing, for example [174]. This problem needs to be addressed hierarchically, because of the problems of scale, encouraging sharing at local, national and international levels. To enable such a progression needs common data standards and it may be possible, for example, to build on the work of the WHO on an International Classification of Diseases (ICD) [209].

Also, there may be merit in working with telecommunications providers to develop generic solutions for tracking and tracing people through their mobile phone location while preserving privacy. The aim would be to provide a platform that could be quickly adapted and adopted by governments or relevant authorities internationally to assist in the control of pandemics. Solving this specific problem is likely to be useful in itself, and give an exemplar of how to learn from big data, and to share data both locally and on a more geographically dispersed basis, while preserving patient anonymity that could usefully be adopted in other contexts.

- **HC5: Learning culture** – given the challenges faced by healthcare there is need to learn from experience both within and out of the domain. There have been previous attempts to draw on experience in other domains, such as aerospace, but with limited success (perhaps the biggest impacts on healthcare from outside have been adoption of safety/assurance cases in some arenas and embracing the Safety-II mindset). However, to be really effective there needs to be a 'learning culture' and a willingness to encourage and foster change at all levels in healthcare – among management, clinical and support staff – and this will require committed leadership.

There are some important initiatives including the adoption of the plan-do-study-act model of improvement [211] and proposals on organisation-wide improvements in healthcare [212]. These approaches have merits but it is unclear if they are sufficient to bring about a learning culture. What is needed is to identify the core principles of safety culture in healthcare and the most effective approaches used to drive improvement looking, among other things, at existing tools and frameworks, and future computational techniques such as those discussed in HC3 and HC4. It will also be necessary to assess whether or not they are consistent with creating the right culture for ensuring safety in increasingly complex systems. In doing this, it will be necessary to work in a coordinated way at different scales – local, such as a hospital, national and international levels. It is likely that this can only be achieved under the auspices of bodies such as the WHO.

These recommendations need to be interpreted and implemented, taking into account the drivers of complexity in healthcare (see Section 5.3). Further, so far as practicable, these activities should be undertaken, or coordinated, on a global basis, most likely by the WHO, in order to ensure their widespread adoption and a reduction in avoidable risks on an international basis. Perhaps the most important issue is to engender more of a 'learning culture' including the willingness to learn from within healthcare and from other domains or industries, as outlined in HC5.

It should be noted that some of the recommendations, for example HC3, are likely to be of interest in other sectors, but it is also important to understand the specific constraints of healthcare, including the variability between patients and the need for a quick response, which means that healthcare may require its own solutions.

It has been reported that NHSX worked with Google and Apple [210] before seeking to develop their own solution for COVID-19, but it may be possible to establish a pan-industry consensus if the developments were carried out under less time pressure.

# D.4 Supply network recommendations

Recommendations for how the safety of complex supply networks can be effectively managed and understood are set out below.

*Governance of supply networks:*

- **SN1:** Regulation needs to be able to look beyond the immediate time to incorporate potential future consequences of supply networks, such as contamination of environment and other downstream safety risks, as frequently the state (future generation of taxpayers) will bear the costs of poor regulation.
- **SN2:** Globalised supply networks cross multiple jurisdictions and therefore require international coordination of standards and regulations, or effective governance and enforcement of local regulations and standards, to ensure supply network safety. Standards and regulations need to be agreed and governed across jurisdictions, or if local regulations and standards deviate there needs to be effective local governance in place.
- **SN3:** Regulation should seek to reduce uncertainty in supply networks perhaps through additional monitoring of inputs to supply networks. This could be done through a standardisation of how data is recorded, with a view to increasing the capacity for tracing of faults or failures, such as contamination of supply networks.
- **SN4:** Methodologies should be developed to assist regulators in the investigation of failures in complex supply networks. The methodologies should seek to avoid the possibility of both premature identification of potential causal links, or premature suspension of investigations when plausible causes are identified.
- **SN5:** Develop regional and international approaches to measuring and detecting leading indication of failure in supply networks, and organisations or monitoring bodies to facilitate this.
- **SN6:** There should be recognition that supply networks are influenced by the social, political, and cultural environments with which they interface.

*Management of supply networks:*

- **SN7:** Information and regulatory guidance needs to be available to assist management with reconciling competing objectives between maintaining the operation of supply networks and maintaining the safety of supply networks; particularly in times of resource scarcity when substitutions of technologies might have to be made to maintain operation.
- **SN8:** The actions of management to balance the potential competing objectives of operation and safety need to be done in the context of the potential for knock-on effects in a highly interdependent supply network. They should be seen as interventions in complex systems that can have consequences in other parts of the system and other times.
- **SN9:** Approaches should be developed to assist management in understanding their organisation's exposure to problems with their supply networks.
- **SN10:** (Senior) management should be aware of their responsibilities in ensuring the safety of their organisation's operations in the long term.
- **SN11:** (Senior) management should be aware that the supply networks that their organisation participates in, or is responsible for, potentially cross multiple cultures.
- **SN12:** (Senior) management should be aware that there might be the need within an organisation for a specific role to be created that takes responsibility for generating a holistic view of the organisation's supply network.

*Task and technology:*

- **SN13:** Technologies could be developed to help lower the uncertainty in supply networks. One area where this could be effectively deployed is tracking and logging of the provenance of resources and materials, supported by a suitable data infrastructure. There is some interesting work in this area that utilises blockchain technology to log the resources, materials and ingredients used in the production of food products including soft drinks [213]. Blockchain technology is also being explored as a means to track the shipments of almonds overseas [214]. Broadly, companies such as IBM are developing tools for monitoring supply chains using blockchain technology [215].
- **SN14:** There is a growing trend in the application of complex networks modelling and simulation techniques being applied to supply network problems, and more work is required to model and simulate supply networks. This should include work to develop generalised approaches to theorising supply networks [81]. There are also specific examples of looking at cascade failures in electricity supply networks [120, 119]. This, or other systematic approaches, for example [216], could also be extended to include efforts to produce systems maps of supply networks, which would be central to efforts to monitor them.

As well as studying supply networks in general, it may be useful to consider them in the context of other domains addressed in the **Safer Complex Systems** study, as there are specific concerns in different domains that might need particular solutions.

E

# References

# References

[1] Simon Burton et al. "Mind the gaps: Assuring the safety of autonomous systems from an engineering, ethical, and legal perspective". In: *Artificial Intelligence* 279 (2020), p. 103201.

[2] Stuart Kauffman. *At Home in the Universe: The Search for the Laws of Self-organization and Complexity*. en. Oxford University Press, Nov. 1996.

[3] L S Schulman and P E Seiden. "Statistical mechanics of a dynamical system based on Conway's game of Life". English. In: *J. Stat. Phys.* 19.3 (Sept. 1978), pp. 293–314.

[4] Peter Erdi. *Complexity Explained*. en. Springer Science & Business Media, Nov. 2007.

[5] Susan Stepney. "Complex Systems for Narrative Theorists". In: *Narrating Complexity*. Ed. by Richard Walsh and Susan Stepney. Cham: Springer International Publishing, 2018, pp. 27–36.

[6] C West Churchman. *Guest editorial: Wicked problems*. 1967.

[7] Marten Scheffer et al. "Early-warning signals for critical transitions". en. In: *Nature* 461.7260 (Sept. 2009), pp. 53–59.

[8] Malcolm Gladwell. *The tipping point: How little things can make a big difference*. Little, Brown, 2006.

[9] Timothy M Lenton. "Early warning of climate tipping points". In: *Nat. Clim. Chang.* 1.4 (July 2011), pp. 201–209.

[10] Marten Scheffer. "Complex systems: Foreseeing tipping points". en. In: *Nature* 467.7314 (Sept. 2010), pp. 411–412.

[11] Ilya Prigogine and Isabelle Stengers. *Order out of chaos: Man's new dialogue with nature*. Verso Books, 2018.

[12] L S D Caves and A T Melo. "(Gardening) Gardening: A relational framework for complex thinking about complex system". In: *Narrating Complexity.* Ed. by R Walsh and S Stepney. London: Springer, 2018.

[13] Ana Teixeira de Melo et al. "Thinking (in) complexity: (In) definitions and (mis)conceptions". In: *Syst. Res. Behav. Sci.* 0.0 (July 2019).

[14] Jens Rasmussen. "Risk management in a dynamic society: a modelling problem". In: *Safety science* 27.2-3 (1997), pp. 183–213.

[15] *New IMarEST President to launch task force to answer the big questions on marine AI*. https:// www.imarest.org/policy-news/ institute-news/ item/5516-new-imarest-president-to-launch-task-force-to-answer- the-big-questions-on-marine-ai. Accessed: 2020-04-13.

[16] Xiaocheng Ge, Richard F Paige, and John A McDermid. "An iterative approach for development of safety-critical software and safety arguments". In: *2010 Agile Conference*. IEEE. 2010, pp. 35–43.

[17] British Standards Institution. *Connected and automated vehicles*. https://www.bsigroup. com/en-GB/cav/. Accessed: 2020-04-13.

[18] Global Mining Guidelines Group. *Guideline for applying functional safety to autonomous systems in mining (Draft in review)*. 2020.

[19] Civil Contingencies Secretariat. "The role of Local Resilience Forums: A reference document". In: *London: Cabinet Office* (2013).

[20] Resilience Engineering Association. *Resilience engineering: Where do I start?* https://www. resilience-engineering-association. org/resources/ where-do-i-start/. Accessed: 2020-04-13.

[21] Alex de Ruijter and Frank Guldenmund. "The bowtie method: A review". In: *Safety science* 88 (2016), pp. 211–218.

[22] Ken Bensinger. *Coronavirus Cases Have Surged, But The US Is Refusing To Take The World's Most Available Masks*. https:// www.buzzfeednews. com/article/ kenbensinger/coronavirus-kn95-masks-us-wont-import- china. Accessed: 2020-04-13.

[23] Geneva Sands and Cristina Alesci. *FDA changes course and allows China's KN95 mask to be used in US*. https://edition.cnn. com/2020/04/03/ politics/fda-china-kn95-us-certain-criteria/ index.html. Accessed: 2020-04-13.

[24] Ana Swanson, Zolan Kanno-Youngs, and Maggie Haberman. *Trump Seeks to Block 3M Mask Exports and Grab Masks From Its Overseas Customers*. https:// www . nytimes. com / 2020 / 04 / 03 / us/ politics/ coronavirus- trump-3m-masks.html. Accessed: 2020-04-13.

[25] BBC News. *Coronavirus: UK failed to stockpile crucial PPE*. https://www. bbc.co.uk/news/ newsbeat-52440641. Accessed: 2020-04-28. 2020.

[26] Andrew Artenstein. *In Pursuit of PPE, New England Journal of Medicine*. https://www.nejm org / doi/full/10.1056 / NEJMc2010025 ? query =featured_coronavirus. 2020.

[27] University of York. *University donates medical equipment and food to support NHS*. https://www. york.ac.uk/news and events/ news/2020/ community/donations-nhs-coronavirus. Accessed: 2020-04-26. 2020.

[28] Cat McDermid. *Intrepid Musicians*. https://vimeo.com /410143485. Accessed: 2020-04-26. 2020.

[29] John McDermid. "System safety engineering: the benefits and practicalities of globalization". In: *Journal of Systems Safety* 47.2 (2011), p. 15.

[30] TUV Rheinland. *Accredited Personnel Certification*. https://www.tuv.com/united-kingdom/en/accredited-diplomas.html. Accessed: 2020-04-13.

[31] Frederick Miller and Judith Katz. *Inclusion Breakthrough: Unleashing the Real Power of Diversity*. en. Berrett-Koehler Publishers, June 2002.

[32] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[33] Brody Huval et al. "An empirical evaluation of deep learning on highway driving". In: *arXiv preprint arXiv:1504.01716* (2015).

[34] Matthieu Komorowski et al. "The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care". In: *Nature medicine* 24.11 (2018), pp. 1716–1720.

[35] David Leslie. "Understanding artificial intelligence ethics and safety". In: *arXiv preprint arXiv:1906.05684* (2019).

[36] Simon Burton, Lydia Gauerhof, and Christian Heinzemann. "Making the case for safety of machine learning in highly automated driving". In: *International Conference on Computer Safety, Reliability, and Security*. Springer. 2017, pp. 5–16.

[37] Richard J Holden et al. "SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients". In: *Ergonomics* 56.11 (2013), pp. 1669–1686.

[38] Andrew Rae, Rob Alexander, and John McDermid. "Fixing the cracks in the crystal ball: A maturity model for quantitative risk assessment". In: *Reliability Engineering & System Safety* 125 (2014), pp. 67–81.

[39] Peter Fenelon and John A McDermid. "An integrated tool set for software safety analysis". In: *Journal of Systems and Software* 21.3 (1993), pp. 279– 290.

[40] C Papadopoulos et al. "Model-based safety assessment for the three stages of refinement of the system development process in ARP4754A". In: *SAE 2011 AeroTech Congress & Exhibition, Toulouse, France*. 2011.

[41] Ewen Denney, Ganesh Pai, and Iain Whiteside. "The role of safety architectures in aviation safety cases". In: *Reliability Engineering & System Safety* 191 (2019), p. 106502.

[42] *SysML Open Source Project What is SysML? Who created SysML?* https://sysml.org. Accessed: 2020-04-13.

[43] Klaus Pohl et al. *Model-based engineering of embedded systems: The SPES 2020 methodology*. Springer Science & Business Media, 2012.

[44] Kester Clegg et al. "Integrating Existing Safety Analyses into SysML". In: *International Symposium on Model-Based Safety and Assessment*. Springer. 2019, pp. 63–77.

[45] Algirdas Avizienis et al. "Basic concepts and taxonomy of dependable and secure computing". In: *IEEE transactions on dependable and secure computing* 1.1 (2004), pp. 11–33.

[46] Bundesministerium für Verkehr und digitale Infrastruktur. *Ethics Commission Report Automated and Connected Driving (in German)*. 2017.

[47] Nidhi Kalra and Susan M Paddock. "Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability?" In: *Transportation Research Part A: Policy and Practice* 94 (2016), pp. 182– 193.

[48] IEC SC 65A. *Functional safety of electrical/electronic/programmable electronic safety-related systems*. Tech. rep. IEC 61508. 3, rue de Varembe´, Case postale 131, CH-1211 Gene`ve 20, Switzerland: The International Electrotechnical Commission, 1998.

[49] Jane Dennett-Thorpe. *Foresight review on the public understanding of risk*. 2017.

[50] Glenn Bruns and Stuart Anderson. "Validating safety models with fault trees". In: *SAFECOMP'93*. Springer, 1993, pp. 21–30.

[51] Robin McDermott, Raymond J Mikulak, and Michael Beauregard. *The basics of FMEA*. SteinerBooks, 1996.

[52] Trevor A Kletz. *HAZOP and HAZAN: identifying and assessing process industry hazards*. IChemE, 1999.

[53] Nancy Leveson. *Engineering a safer world: Systems thinking applied to safety*. MIT press, 2011.

[54] Erik Hollnagel. *Safety-II in practice: developing the resilience potentials*. Taylor & Francis, 2017.

[55] Erik Hollnagel, David D Woods, and Nancy Leveson. *Resilience engineering: Concepts and precepts*. Ashgate Publishing, Ltd., 2006.

[56] Erik Hollnagel. *FRAM, the functional resonance analysis method: modelling complex socio-technical systems*. Ashgate Publishing, Ltd., 2012.

[57] Roman Gansch and Ahmed Adee. "System Theoretic View on Uncertainties". In: *DATE 2020*. 2020.

[58] Charles Perrow. *Normal accidents: Living with high risk technologies-Updated edition*. Princeton university press, 2011.

[59] International Civil Aviation Organization. *State of Global Aviation Safety ICAO Safety Report 2019 Edition*.

[60] International Civil Aviation Organization. *10004: Global Aviation Safety Plan 2020-2022*.

[61] Naaman Zhou. *Volvo admits its self-driving cars are confused by kangaroos*. https://www.theguardian.com/technology/2017/jul/01/volvo-admits- its-self-driving-cars-are-confused-by-kangaroos. Accessed: 202004-29.

[62] Reuters. *Transport safety body rules safeguards 'were lacking' in deadly Tesla crash*. https://www.theguardian.com/technology/2017/sep/12/tesla-crash-joshua-brown-safety-self-driving-cars. Accessed: 2020-04-29.

[63] *European Commission: Guidelines on the exemption procedure for the EU approval of automated vehicles. Version 4.1.* https://ec.europa.eu/growth/content/guidelines-exemption-procedure-eu-approval-automated-vehicles_en. 2019.

[64] *Automated Driving Systems 2.0, A Vision for Safety*. https://www.nhtsa. gov/sites/nhtsa.dot .gov/file/documents/13069aads2.0_090617_ v9a_tag.pdf. 2017.

[65] Economic Commission for Europe, Inland Transport Committee. *Revised Framework document on automated/autonomous vehicles*. Tech. rep. ECE/TRANS/WP.29/2019/34/Rev.2. United Nations, 2019.

[66] International Organization for Standardization. *Road vehicles — Safety of the intended functionality*. Tech. rep. ISO/PAS 21448:2019. ISO, 2019.

[67] Underwriters Laboratories. *Standard for Safety for the Evaluation of Autonomous Products*. Tech. rep. ANSI/UL 4600. 2019.

[68] PJ Clarkson et al. *Engineering better care: a systems approach to health and care design and continuous improvement*. 2017.

[69] *How does UK healthcare spending compare with other countries?* https://www.ons.gov.uk / people population and community /healthandsocialcare/healthcaresystem / articles/ doesukhealthcare spendingcomparewith othercountries/ 2019-08-29. Accessed: 2020-04-13.

[70] *Current health expenditure (% of GDP)*. https://data. worldbank.org/indicator/SH.XPD.CHEX.GD.ZS. Accessed: 2020-04-13.

[71] U.S. Food & Drug Administration and others. "Infusion pumps total product life cycle: Guidance for industry and FDA staff". In: *Food and Drug Administration Standard* (2014), pp. 910–766.

[72] Jeffrey Braithwaite, Robert L Wears, and Erik Hollnagel. "Resilient health care: turning patient safety on its head". In: *International Journal for Quality in Health Care* 27.5 (2015), pp. 418–420.

[73] Erik Hollnagel and Jeffrey Braithwaite. *Resilient health care*. CRC Press, 2019.

[74] NHS Improvement. "Revised Never Events policy and framework and never events list 2018". In: *London, UK: NHS Improvement* (2018).

[75] NHS Improvement. "Never events list 2018". In: *London: NHS Improvement* (2018).

[76] NHS Improvement. "NRLS national patient safety incident reports: commentary". In: (2020).

[77] Jose M Valderas et al. "Defining comorbidity: implications for understanding health and health services". In: *The Annals of Family Medicine* 7.4 (2009), pp. 357–363.

[78] Daniel Shu Wei Ting et al. "Artificial intelligence and deep learning in ophthalmology". In: *British Journal of Ophthalmology* 103.2 (2019), pp. 167– 175.

[79] Izet Masic, Milan Miokovic, and Belma Muhamedagic. "Evidence based medicine – new approaches and challenges". In: *Acta Informatica Medica* 16.4 (2008), p. 219.

[80] Eduardo Hariton and Joseph J Locascio. "Randomised controlled trials—The gold standard for effectiveness research". In: *BJOG: an international journal of obstetrics and gynaecology* 125.13 (2018), p. 1716.

[81] Edward JS Hearnshaw and Mark MJ Wilson. "A complex network approach to supply chain network theory". In: *International Journal of Operations & Production Management* (2013).

[82] Andrew G Haldane and Robert M May. "Systemic risk in banking ecosystems." In: *Nature* 469.7330 (Jan. 2011), pp. 351–5. ISSN: 1476-4687. DOI: 10 . 1038 / nature09659. URL: http:// www. ncbi. nlm. nih. gov/pubmed/ 21248842.

[83] Sergey V Buldyrev et al. "Catastrophic cascade of failures in interdependent networks". In: *Nature* 464.7291 (Apr. 2010), pp. 1025–1028. ISSN: 00280836. URL: http://dx.doi.org/10.1038/nature08932%20https://www.nature.com/nature/journal/v464/n7291/full/nature08932.html.

[84] *Supply chains have been upended. Here's how to make them more resilient*. https://www.weforum.org/agenda/2020/04/supply-chains-resilient- covid-19/. Accessed: 2020-6-11.

[85] Philip Garnett, Bob Doherty, and Tony Heron. "Vulnerability of the United Kingdom's food supply chains exposed by COVID-19". In: *Nature Food* (June 2020).

[86] "Trade Secrets: Supply Chain Disruption". In: *Financial Times* (May 2020).

[87] Haralambos Sarimveis et al. "Dynamic modeling and control of supply chain systems: A review". In: *Comput. Oper. Res.* 35.11 (Nov. 2008), pp. 3530–3561.

[88] Detlef Zerfowski and Andreas Lock. "Functional architecture and E/E-Architecture– A challenge for the automotive industry". In: *19. Internationales Stuttgarter Symposium*. Springer. 2019, pp. 909–920.

[89] Roger Kemp. *Living without electricity One city's experience of coping with loss of power*. 2016.

[90] NASA. *Hidden Hazards*. https://nsc.nasa.gov/resources/case-studies/detail/hidden-hazards. Accessed: 2020-04-16. 2017.

[91] International Labour Organization. *ILO Standards and COVID-19 (coronavirus), version 1.2*. 2020.

[92] Boeing Commercial Airplanes. "Statistical Summary of Commercial Jet Airplane Accidents–Worldwide Operations, 1959–2018". In: *Aviation Safety, Seattle* (2019).

[93] Highways England. *Smart Motorways*. https://highwaysengland.co.uk/programmes/smart-motorways/. Accessed: 2020-04-29.

[94] EuroNCAP. *The European New Car Assessment Programme*.

[95] *How is pollution is destroying our health, author=WHO*. https://www.who.int/airpollution/newsand-events/how-air-pollution-is-destroying-our-health. Accessed: 2020-07-02. 2020.

[96] OECD/International Transport Forum. *Road Safety Annual Report 2019*. https://www.itfoecd.org/sites/default/files/docs/irtadroad- safety-annual-report-2019.pdf. Accessed: 2020-04-16. 2019.

[97] World Health Organization. *Global Status Report on Road Safety 2015*. https:// www . who . int/ violence_ injury_ prevention/ road _ safety_ status/2015/en/. Accessed: 2020-04-16. 2015.

[98] John A McDermid and AJ Rae. "How did systems get so safe without adequate analysis methods?, in Proc., 9th IET International Conference on System Safety and Cyber Security". In: (2014).

[99] Fredrik Asplund et al. "Rapid Integration of CPS Security and Safety". In: *IEEE Embedded Systems Letters* 11.4 (2018), pp. 111–114.

[100] Simon Burton et al. "Automotive functional safety= safety+ security". In: *Proceedings of the First International Conference on Security of Internet of Things*. 2012, pp. 150–159.

[101] Jane L Fenn et al. "The who, where, how, why and when of modular and incremental certification". In: *2nd IET International Conference on System Safety*. IET. 2007, pp. 135–140.

[102] Ewen Denney, Ganesh Pai, and Ibrahim Habli. "Dynamic safety cases for through-life safety assurance". In: *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*. Vol. 2. IEEE. 2015, pp. 587–590.

[103] Care Quality Commission and Medical and Healthcare products Regulatory Agency. *Using machine learning in diagnostic services: A report with recommendations from CQC's regulatory sandbox*. 2020.

[104] Sidney Dekker. *Just culture: Balancing safety and accountability*. Ashgate Publishing, Ltd., 2012.

[105] UK Government. *HSWA*. http://www.legislation.gov.uk/ukpga/1974/37/contents. Accessed: 2020-04-16. 1974 (First Published).

[106] Spike WS Lee, Julie Y Huang, and Norbert Schwarz. "Risk Overgeneralization in Times of a Contagious Disease Threat". In: *Frontiers in Psychology* 11 (2020), p. 1392.

[107] Toby Wise et al. *Changes in risk perception and protective behavior during the first week of the COVID-19 pandemic in the United States*. Mar. 2020. DOI: 10.31234/osf.io/dz428.

[108] John D Lee and Katrina A See. "Trust in automation: Designing for appropriate reliance". In: *Human factors* 46.1 (2004), pp. 50–80.

[109] James Zou and Londa Schiebinger. *AI can be sexist and racist—it's time to make it fair*. 2018.

[110] The Guardian. *Failure to publish data on BAME deaths could put more lives at risk, MPs warn*. https://www. theguardian. com/world/2020 /apr/16/ dataon bame deaths fromcovid19 mustbepublished- politicians-warn. Accessed: 2020-04-20. 2020.

[111] Srini Sundaram et al. "Aircraft engine health monitoring using density modelling and extreme value statistics". In: *Proceedings of the 6th International Conference on Condition Monitoring and Machine Failure Prevention Technologies*. Vol. 3. 2009.

[112] Rolls-Royce. *Rolls-Royce IntelligentEngine vision makes rapid progress*. https:// www. rollsroyce. com/ media/ pressreleases/ 2018 / 16 07 2018 - rr-intelligentengine-vision-makes-rapid-progress.aspx. Accessed: 2020-04-19. 2018.

[113] Slingshot. *Digital Twins for a Breathing City*. https://www. slingshotsimulations. co.uk. Accessed: 2020-06-25. 2020.

[114] Lloyd's Register Foundation. *Unlocking the potential of health and safety data*. https://www.lr.org/en-gb/latest-news/unlocking-h-s-data/. Accessed: 2020-04-19. 2020.

[115] John Alexander McDermid, Yan Jia, and Ibrahim Habli. "Towards a Framework for Safety Assurance of Autonomous Systems". In: *Artificial Intelligence Safety 2019*. CEUR Workshop Proceedings. 2019, pp. 1–7.

[116] Nicola Paltrinieri, Louise Comfort, and Genserik Reniers. "Learning about risk: Machine learning for risk assessment". In: *Safety science* 118 (2019), pp. 475–486.

[117] Katherine Haver. "Haiti Earthquake Response: Mapping and analysis of gaps and duplications in evaluations". In: *Haiti Earthquake Response: Mapping and analysis of gaps and duplications in evaluations*. Active Learning Network for Accountability and Performance in humanitarian . . ., 2011.

[118] Cheng Zhou et al. "Characterizing time series of near-miss accidents in metro construction via complex network theory". In: *Saf. Sci.* 98 (Oct. 2017), pp. 145–158.

[119] Paolo Crucitti, Vito Latora, and Massimo Marchiori. "Model for cascading failures in complex networks". en. In: *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 69.4 Pt 2 (Apr. 2004), p. 045104.

[120] Paolo Crucitti, Vito Latora, and Massimo Marchiori. "A topological analysis of the Italian electric power grid". In: *Physica A: Statistical Mechanics and its Applications* 338.1 (July 2004), pp. 92–97.

[121] A Dwivedi, X Yu, and P Sokolowski. "Identifying vulnerable lines in a power network using complex network theory". In: *2009 IEEE International Symposium on Industrial Electronics*. July 2009, pp. 18–23.

[122] Geir Kjetil Hanssen, Tor Sta˚ lhane, and Thor Myklebust. *SafeScrum* R *-Agile Development of Safety-Critical Software*. Springer, 2018.

[123] YongBin Zhou and DengGuo Feng. "Side-Channel Attacks: Ten Years After Its Publication and the Impacts on Cryptographic Module Security Testing." In: *IACR Cryptology ePrint Archive* 2005.388 (2005).

[124] Ivan Evtimov et al. "Robust physical-world attacks on deep learning models". In: *arXiv preprint arXiv:1707.08945* (2017).

[125] J McDermid. "Playing catch-up: The fate of safety engineering". In: *Developments in System Safety Engineering, Proceedings of the Twenty-fifth SafetyCritical Systems Symposium, Bristol, UK, ISBN*. 2017, pp. 978–1540796288.

[126] Andrew H Van de Ven et al. *Engaged scholarship: A guide for organizational and social research*. Oxford University Press on Demand, 2007.

[127] Thomas S Kuhn. *The structure of scientific revolutions*. University of Chicago press, 2012.

[128] Karl Popper. *The logic of scientific discovery*. Routledge, 2005.

[129] Imre Lakatos. "The role of crucial experiments in science". In: *Studies in History and Philosophy of Science Part A* 4.4 (1974), pp. 309–325.

[130] Horst W J Rittel and Melvin M Webber. "Dilemmas in a general theory of planning". In: *Policy Sci.* 4.2 (June 1973), pp. 155–169.

[131] Ludwig Von Bertalanffy. "General system theory". In: *New York* 41973 (1968), p. 40.

[132] William D Ross et al. *Aristotle's metaphysics*. Oxford University Press, 1925.

[133] Matthew Syed. *Rebel Ideas: The Power of Diverse Thinking*. Hachette UK, 2019.

[134] Defence Safety Authority. *Service Inquiry, Loss of Watchkeeper (WK043) Unmanned Air Vehicle over Cardigan Bay in West Wales 24 Mar 17 (DSA/DAIB/17/006)*. 2019.

[135] Uniting Aviation. *High-altitude long-endurance aircraft are seeking new operational flight levels*. https://unitingaviation.com/news/safety/high-fliers-high-altitude-long-endurance-aircraft-are-seeking-new-operational-flight-levels/ Accessed: 2020-04-13.

[136] House Committee on Transportation and Infrastructure. *The Boeing 737 MAX Aircraft: Costs, Consequences, and Lessons from its Design, Development, and Certification -Preliminary Investigative Findings*. 2020.

[137] CH Cave. "An independent review into the broader issues surrounding the loss of the RAF Nimrod MR2 Aircraft XV230 in Afghanistan in 2006". In: *The Stationary Office, Tech. Rep* (2006).

[138] J Hackitt. "Building a safer future—Independent review of building regulations and fire safety: final report". In: *UK Gov* (2018).

[139] Main Commission and others. "Report on the accident to Airbus A320-211 Aircraft in Warsaw on 14 September 1993". In: *Rapport technique, Aircraft Accident Investigation Warsaw, Poland* (1994), p. 32.

[140] Agile Alliance. *C12 Principles Behind the Agile Manifesto*.

[141] "The risks of LSCITS: the odds are stacked against us". In:

[142] IET and HSE. *Code of Practice: Competence for Safety Related Systems Practitioners*. IET Standards, 2016.

[143] Georgios Kavallieratos, Sokratis Katsikas, and Vasileios Gkioulos. "Cybersecurity and Safety Co-Engineering of Cyberphysical Systems—A Comprehensive Survey". In: *Future Internet* 12.4 (2020), p. 65.

[144] Kamrul Hasan et al. "Towards Optimal Cyber Defense Remediation in Cyber Physical Systems by Balancing Operational Resilience and Strategic Risk". In: *MILCOM 2019-2019 IEEE Military Communications Conference (MILCOM)*. IEEE. 2019, pp. 1–8.

[145] Richard Hawkins and John McDermid. *Safety assurance of autonomy to support the Fourth Industrial Revolution*. 2019.

[146] Xueyi Zou, Rob Alexander, and John McDermid. "Safety validation of sense and avoid algorithms using simulation and evolutionary search". In: *International Conference on Computer Safety, Reliability, and Security*. Springer. 2014, pp. 33–48.

[147] ICGAI. *International Congress for the Governance of AI*. https:// icgai. org. Accessed: 2020-04-13.

[148] *Defence Committee launches Sub-Committee on the security of 5G*. https: //www.parliament.uk/ business/committees/committees-a-z/commons- select / defence committee / news parliament 2017 / security 5g - inquiry-launch-19-21/. Accessed: 2020-04-13.

[149] Nicole A Cooke. *Fake news and alternative facts: Information literacy in a post-truth era*. American Library Association, 2018.

[150] BBC. *Mast fire probe amid 5G coronavirus claims*. https:// www.bbc.co. uk/news/uk-england-52164358. Accessed: 2020-04-13.

[151] Richard M J Bohmer et al. "How Hospitals Can Manage Supply Shortages as Demand Surges". In: *Harvard Business Review* (Apr. 2020).

[152] "Supply chain disruption: sunken ambitions". In: *Financial Times* (Nov. 2011).

[153] Stuart Pugh. *Total design: integrated methods for successful product engineering*. Addison-Wesley, 1991.

[154] JI Goatham. "Materials problems in the design of compressor blades with fibrous composites". In: *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 319.1536 (1970), pp. 45–57.

[155] Henry Petroski. *To engineer is human: The role of failure in successful design*. St Martins Press, 1985.

[156] Bob Wallace. "AT&T outage locks up user nets, sends workers home". In: *Network World, Vol. 7, No. 4*.

[157] Robert Walmsley et al. *NATS System Failure 12 December 2014 – Final Report Independent Enquiry*. May 2015.

[158] Prepared by the Democratic Staff of the House Committee on Transportation and Infrastructure for Chair Peter A. DeFazio, Sub-committee on Aviation Chair Rick Larsen, and Members of the Committee. *The Boeing 737 MAX Aircraft: Costs, Consequences, and Lessons from its Design, Development, and Certification. Preliminary Investigative Findings*. Mar. 2020.

[159] Booz Allen Hamilton. *Executive Briefing – Urban Air Mobility (UAM) Market Study*. Oct. 2018.

[160] Presidential Commission on the Space Shuttle Challenger Accident. *Report of the Presidential Commission on the Space Shuttle Challenger Accident*. June 1986.

[161] Investigation Commission. *Final Report concerning the accident which occurred on June 26th 1988 at Mulhouse-Habsheim (68) to the Airbus A 320, registered F-GFKC*. Nov. 1989.

[162] German Federal Bureau of Aircraft Accidents Investigation. *Investigation Report AX001-1-2/02*. May 2004.

[163] Alvin Wilby. *Private Communication*. 2019.

[164] Defence Safety Authority. "Service Inquiry, Loss of Watchkeeper (WK043) Unmanned Air Vehicle over Cardigan Bay in West Wales 24 Mar 17". In: (2019).

[165] AJ Arlow, CJ Duffy, and JA McDermid. "Safety specification of the active traffic management control system for English motorways". In: (2006).

[166] European Transport Safety Council. *UK government puts new "smart" motorways on hold*. https:// etsc. eu/ ukgovernmentputsnewsmart-motorways-on-hold/. Accessed: 2020-04-29.

[167] National Transportation Safety Board. "Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018". In: (2019).

[168] National Transportation Safety Board. "Collision Between a Car Operating With Automated Vehicle Control Systems and a Tractor-Semitrailer Truck Near Williston, Florida". In: (2017).

[169] Charlie Miller and Chris Valasek. "Remote exploitation of an unaltered passenger vehicle". In: *Black Hat USA* 2015 (2015), p. 91.

[170] Health Service Executive. *Guideline for the Systems Analysis Investigation of Incidents and Complaints, Revision 1.1*. 2012.

[171] Sabaratnam Arulkumaran. "Investigation of Incident 50278 from time of patient's self referral to hospital on the 21st of October 2012 to the patient's death on the 28th of October, 2012". In: (2013).

[172] Irish Medical Council. *Guide to professional conduct and ethics for registered medical practitioners*. 2009.

[173] Worldometer. *Confirmed Cases and Deaths by Country, Territory, or Conveyance*. https:// www.worldometers. info/ coronavirus/. Accessed: 2020-03-31.

[174] Zhejiang University School of Medicine. *Handbook of COVID-19 Prevention and Treatment*. 2020.

[175] NM Ferguson, D Laydon, G Nedjati-Gilani, et al. *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imperial College COVID-19 Response Team*. 2020.

[176] Al Jazeera. *Timeline: How the new coronavirus spread*. https:// www.aljazeera.com/news/2020/01/ timeline china coronavirusspread-200126061554884.html. Accessed: 2020-04-13.

[177] David Batty. "E coli outbreak leads to French ban on seeds from British firm". In: *The Guardian* (June 2011).

[178] Ian Sample. "E coli outbreak: German organic farm officially identified". In: *The Guardian* (June 2011).

[179] "Escherichia coli (E. coli) outbreak in United Kingdom". en. In: *World Health Organization* (July 2016).

[180] James Meikle. "Mixed salad leaves linked to E coli outbreak that has killed two in UK". In: *The Guardian* (July 2016).

[181] Public Health England. *E. coli O157 national outbreak update*. https:// www.gov.uk/government/ news/update-as-e-coli-o157- investigation- continues. Accessed: 2020-04-02. July 2016.

[182] "EHEC outbreak: 9 European countries report cases of haemolytic uraemic syndrome and enterohaemorrhagic E. coli infections". en. In: *World Health Organization* (May 2011).

[183] Mark K Johansen and Magda Osman. "Coincidence judgment in causal reasoning: How coincidental is this?" en. In: *Cogn. Psychol.* 120 (Aug. 2020), p. 101290.

[184] Magda Osman. "Controlling uncertainty: a review of human behavior in complex dynamic environments". en. In: *Psychol. Bull.* 136.1 (Jan. 2010), pp. 65– 86.

[185] Zhanyun Wang et al. "A Never-Ending Story of Perand Polyfluoroalkyl Substances (PFASs)?" en. In: *Environ. Sci. Technol.* 51.5 (Mar. 2017), pp. 2508– 2518.

[186] Stockholm Convention. *Listing of POPs in the Stockholm Convention*. http: //www.pops.int/ TheConvention/ThePOPs/AllPOPs/ tabid/2509/Default. aspx. Accessed: 2020-4-8.

[187] Marianne Haukås et al. "Bioaccumulation of per and polyfluorinated alkyl substances (PFAS) in selected species from the Barents Sea food web". en. In: *Environ. Pollut.* 148.1 (July 2007), pp. 360–371.

[188] Marko Filipovic et al. "Historical usage of aqueous film forming foam: a case study of the widespread distribution of perfluoroalkyl acids from a military airport to groundwater, lakes, soils and fish". en. In: *Chemosphere* 129 (June 2015), pp. 39–45.

[189] Anna Rotander et al. "Elevated levels of PFOS and PFHxS in firefighters exposed to aqueous film forming foam (AFFF)". en. In: *Environ. Int.* 82 (Sept. 2015), pp. 28–34.

[190] Jenny Bytingsvik et al. "Perfluoroalkyl substances in polar bear mother–cub pairs: A comparative study based on plasma levels from 1998 and 2008". In: *Environ. Int.* 49 (Nov. 2012), pp. 92–99.

[191] *Is the downfall of AFFF a precursor to long term environmental liabilities?* https:// www.internationalairportreview. com/article/98795/fire- fighting-foam-chemicals-water/. Accessed: 2020-4-17.

[192] The State of Queensland, Australia. *Firefighting Foam Operational Policy— Overview*. https://www.qld.gov.au/ environment/pollution/ management/ disasters/ investigationpfas/firefightingfoam/ policyoverview. Accessed: 2020-4-17.

[193] Independent Investigation Board. "Train derailment at Hatfield: A final report by the independent investigation board". In: *Report. Office of Rail Regulation* (2006).

[194] EMSD. *Investigation Report on Incident of the New Signalling System Testing on MTR Tsuen Wan Line*. EMSD, 2019.

[195] Deane McNulty and Philip Rielly. "King's Cross fire in the London Underground November 18, 1987: a report". In: (1992).

[196] Lord W Douglas Cullen. "The public inquiry into the Piper Alpha disaster". In: *Drilling Contractor;(United States)* 49.4 (1993).

[197] Republic of the Marshall Islands Office of the Maritime Administrator. "Deepwater Horizon Marine Casualty Investigation Report". In: (2001).

[198] Scott A Snook. *Friendly fire: The accidental shootdown of US Black Hawks over northern Iraq*. Princeton university press, 2002.

[199] Charles Haddon-Cave QC. "The Nimrod Review". In: *An Independent Review into the Broader Issues Surrounding the Loss of the RAF Nimrod MR2 Aircraft XV230 in Afghanistan in 2006* (2009).

[200] Martina K Linnenluecke and Andrew Griffiths. "The 2009 Victorian bushfires: A multilevel perspective on organizational risk and resilience". In: *Organization & Environment* 26.4 (2013), pp. 386–411.

[201] U.S. House of Representatives. "Select Bipartisan Committee to Investigate the Preparation for and Response to Hurricane Katrina". In: (2005).

[202] Samsung. *Samsung Global Privacy Policy SmartTV Supplement*. https://www.samsung.com/uk/info/privacy-SmartTV/?CID=AFL-hq-mul-0813- 11000170. Accessed: 2020-05-17. 2020.

[203] Simon Burton et al. "Confidence arguments for evidence of performance in machine learning for highly automated driving functions". In: *International Conference on Computer Safety, Reliability, and Security*. Springer. 2019, pp. 365–377.

[204] Peter Spurgeon et al. *Building Safer Healthcare Systems: A Proactive, Risk Based Approach to Improving Patient Safety*. Springer Nature, 2019.

[205] Yan Jia. "Improving medication safety using machine learning". In: *AIME 2019* (2019).

[206] Chartered Institute of Ergonomics & Human Factors. *Guidance to help design effective and usable work procedures for health and social care teams*. Tech. rep. CHIEF.

[207] Chiara Picardi et al. "Assurance Argument Patterns and Processes for Machine Learning in Safety-Related Systems". In: *Proceedings of the Workshop on Artificial Intelligence Safety (SafeAI 2020)*. CEUR Workshop Proceedings. 2020, pp. 23–30.

[208] U.S. Food & Drug Administration. *Artificial Intelligence and Machine Learning in Software as a Medical Device*. https:// www . fda. gov / medical- devices/ software medicaldevicesamd/ artificialintelligence-andmachinelearningsoftwaremedicaldevice. Accessed: 2020-0413.

[209] WHO. *International Classification of Diseases (ICD)*.

[210] Rowena Mason. *UK app to track coronavirus spread to be launched*. https: // www. theguardian. com/ politics/ 2020 / apr/ 12 / ukapptotrack- coronavirus-spread-to-be-launched. Accessed: 2020-04-13.

[211] NHS Improvement. "Plan, Do, Study, Act (PDSA) cycles and the model for improvement". In: *London: NHS* (2018).

[212] Bryan Jones, Tim Horton, and Will Warburton. "The improvement journey". In: *Why organisation wide improvement in health care matters, and how to get started* (2019).

[213] *Almond*. https://almond.org/. Accessed: 2020-4-20.

[214] *Almond Supply Chain Tracked with Blockchain*. https:// www.foodlogistics. com/ technology/news/21015561/ almond supplychaintrackedwith-blockchain. Accessed: 2020-4-20. July 2018.

[215] *Blockchain for Supply Chain IBM Blockchain*. https://www.ibm. com/uk- en/blockchain/industries/ supply-chain. Accessed: 2020-5-14.

[216] Shi-Jie (Gary) Chen and Enzhen Huang. "A systematic approach for supply chain improvement using design structure matrix". In: *J. Intell. Manuf.* 18.2 (Apr. 2007), pp. 285–299.

# Safer Complex Systems

## An initial framework for understanding and improving the safety of complex, interconnected systems in a rapidly changing and uncertain world