Identifying and Addressing Challenges for Safe and Secure Complex Systems

JOHN MCDERMID, 12TH MAY 2023

Agenda

- Safer Complex Systems Study
- Defining Complex Systems
- Challenges
- A Framework for Managing Safety
- Some Examples
- Safety and Security
- Some Principles
- Conclusions

Engineering X

Safer Complex Systems

An Initial Framework

Royal Academy of Engineering

Lloyd's Register Foundation

Safer Complex Systems

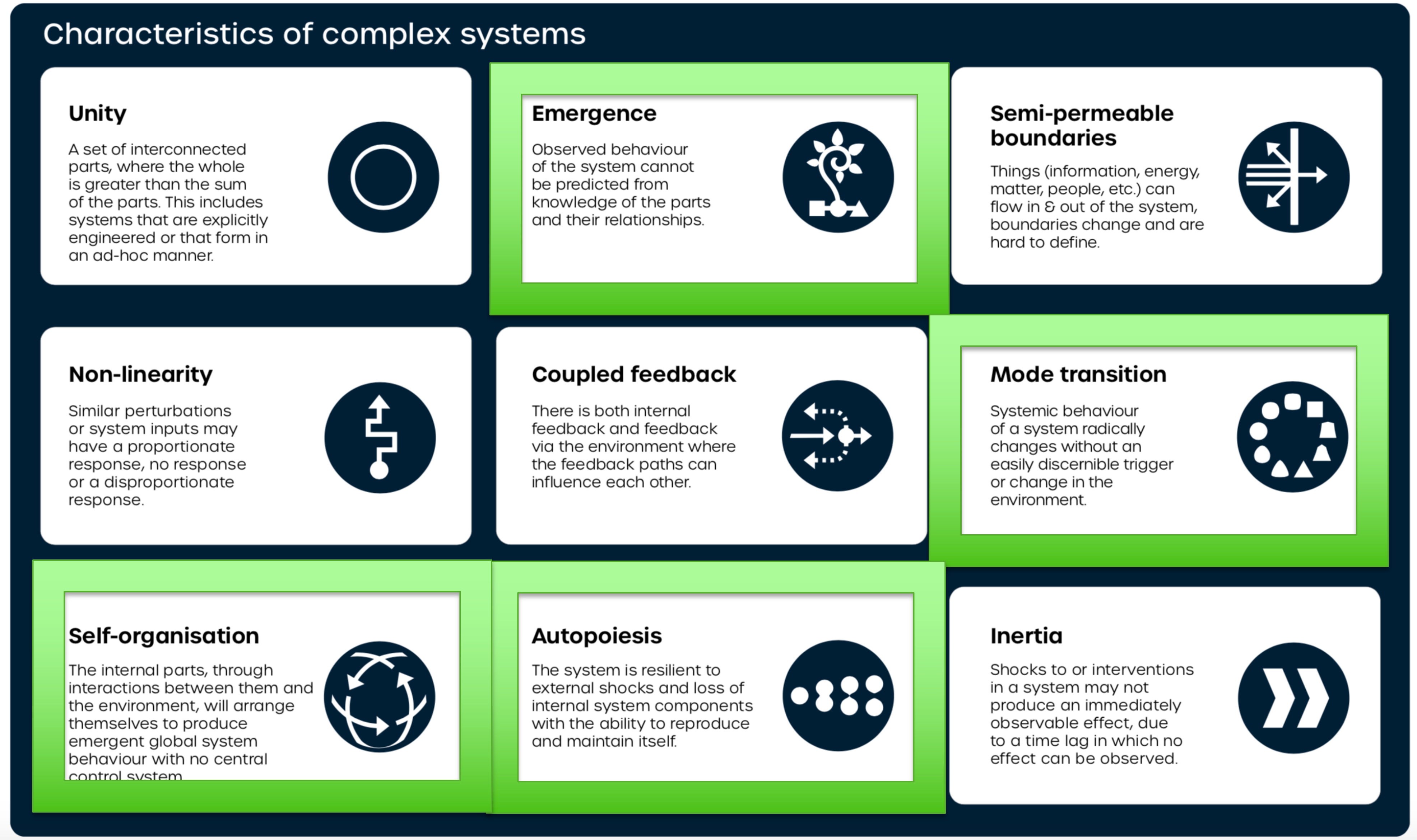
Project Aims

- 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
- To develop an 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)
- To outline emerging challenges and opportunities with significant disruptive potential (negative or positive) with regards to the safety of complex systems

https://raeng.org.uk/media/4wxiazh3/engineering-x-safer-complex-systems-an-initial-framework-report-v22.pdf

Defining Complexity

Characteristics



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 & 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.

Challenges

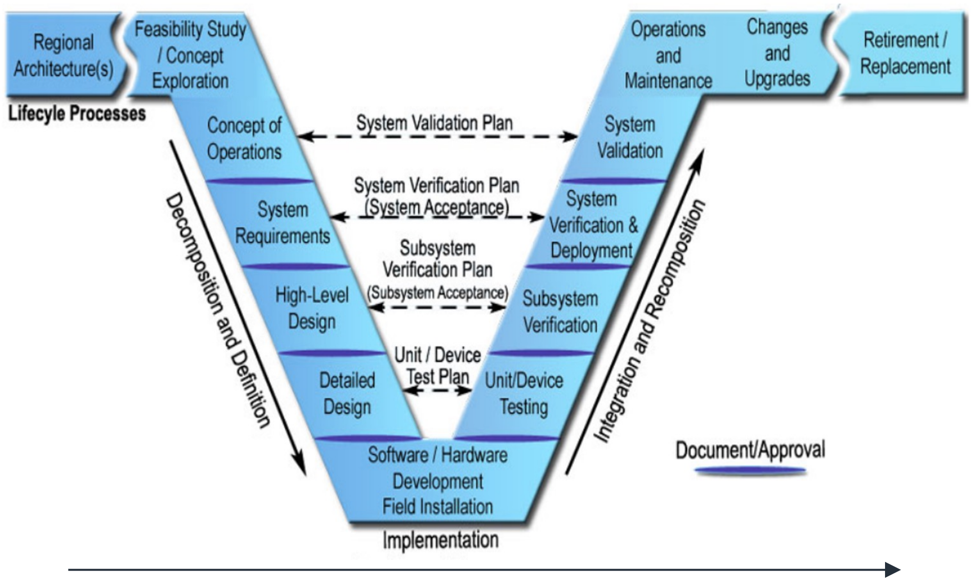
Emergence and More

- Often "emergence" is viewed as the defining characteristic of complex systems
  - As opposed to merely "complicated" systems
  - But other characteristics, and really a "multi-dimensional spectrum"
- Also need to consider distinction between
  - Systems, including systems of systems, designed as a whole (with a "controlling mind"), e.g. a car, an aircraft, commercial air traffic?
  - More "ad hoc" systems, not designed as a whole, e.g. road traffic (there are partial controlling minds, but no overall control, such as the introduction of partially autonomous vehicles in the USA)

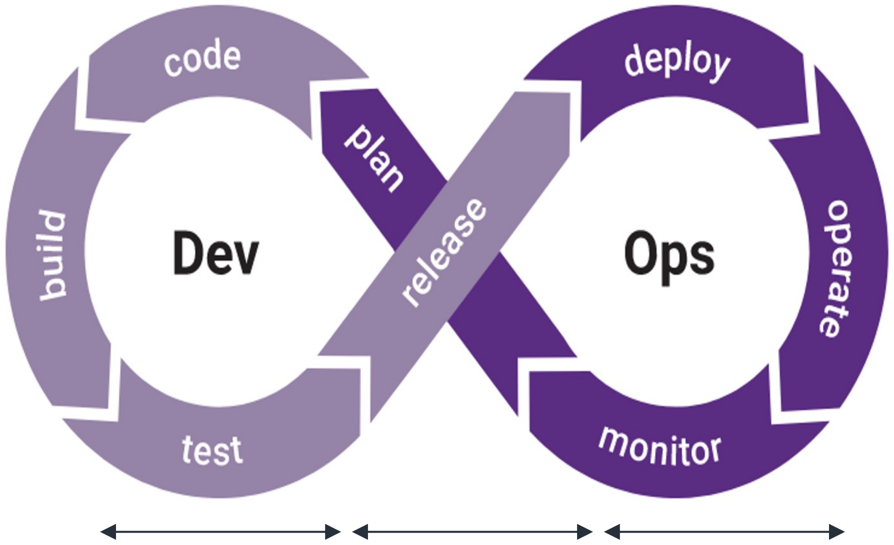
Challenges

Dynamics

- Development processes mean that systems evolve very quickly
  - Moving from “V” to DevOps
  - Challenges both safety and security processes



Six Months to Two Years

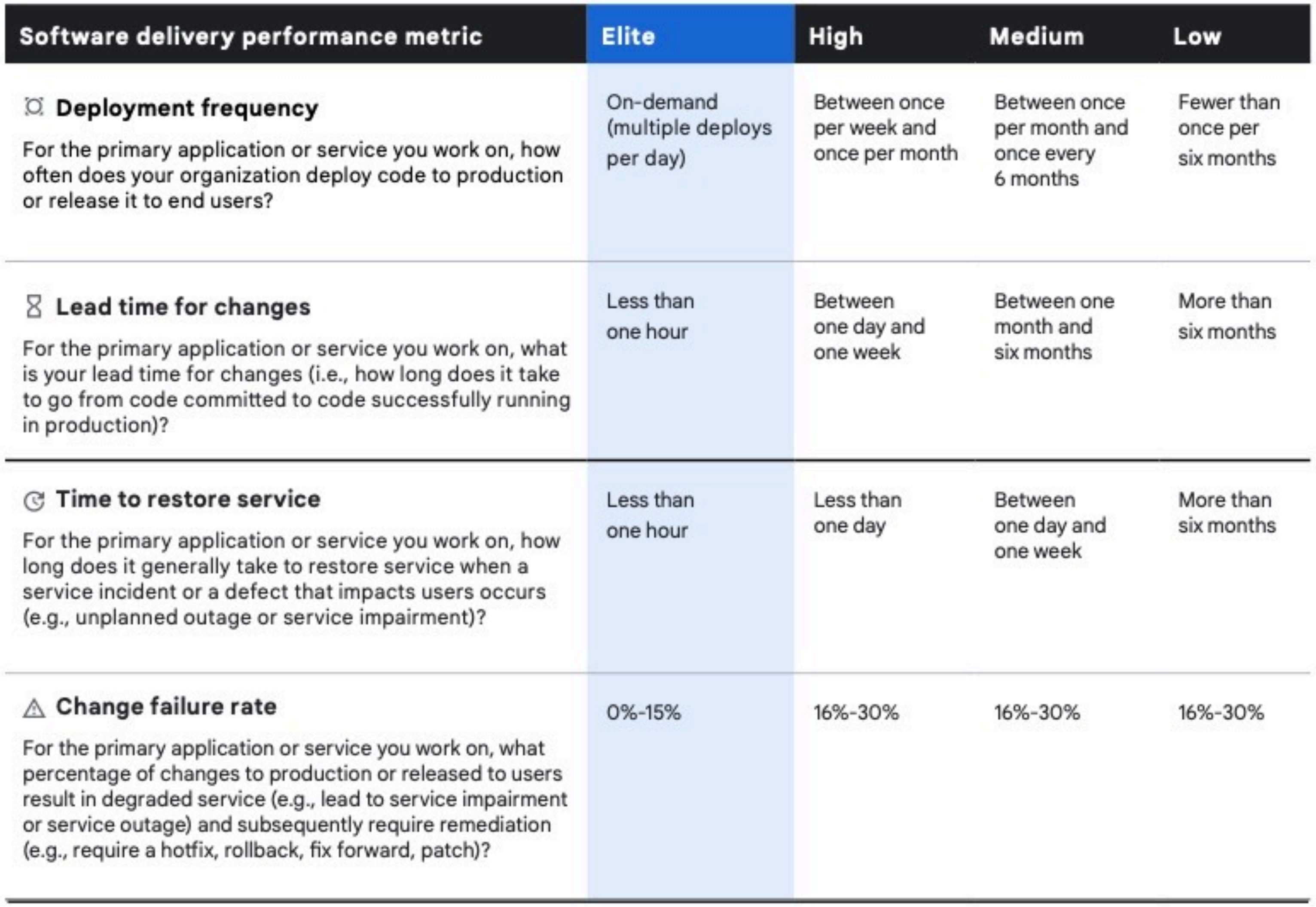


Days Hours Seconds

Challenges

Failure and Repair

- High rate of change brings high failure rate
  - Even for so-called elite organisations
  - Also very hard to analyse due to volatility, including “fixes to fixes”



Software delivery performance metric	Elite	High	Medium	Low
Deployment frequency For the primary application or service you work on, how often does your organization deploy code to production or release it to end users?	On-demand (multiple deploys per day)	Between once per week and once per month	Between once per month and once every 6 months	Fewer than once per six months
Lead time for changes For the primary application or service you work on, what is your lead time for changes (i.e., how long does it take to go from code committed to code successfully running in production)?	Less than one hour	Between one day and one week	Between one month and six months	More than six months
Time to restore service For the primary application or service you work on, how long does it generally take to restore service when a service incident or a defect that impacts users occurs (e.g., unplanned outage or service impairment)?	Less than one hour	Less than one day	Between one day and one week	More than six months
Change failure rate For the primary application or service you work on, what percentage of changes to production or released to users result in degraded service (e.g., lead to service impairment or service outage) and subsequently require remediation (e.g., require a hotfix, rollback, fix forward, patch)?	0%-15%	16%-30%	16%-30%	16%-30%

UNIVERSITY of York

Agenda

- Safer Complex Systems Study
- Defining Complex Systems
- Challenges
- A Framework for Managing Safety
- Some Examples
- Safety and Security
- Some Principles
- Conclusions

EngineeringX

Safer Complex Systems
An Initial Framework

Royal Academy of Engineering

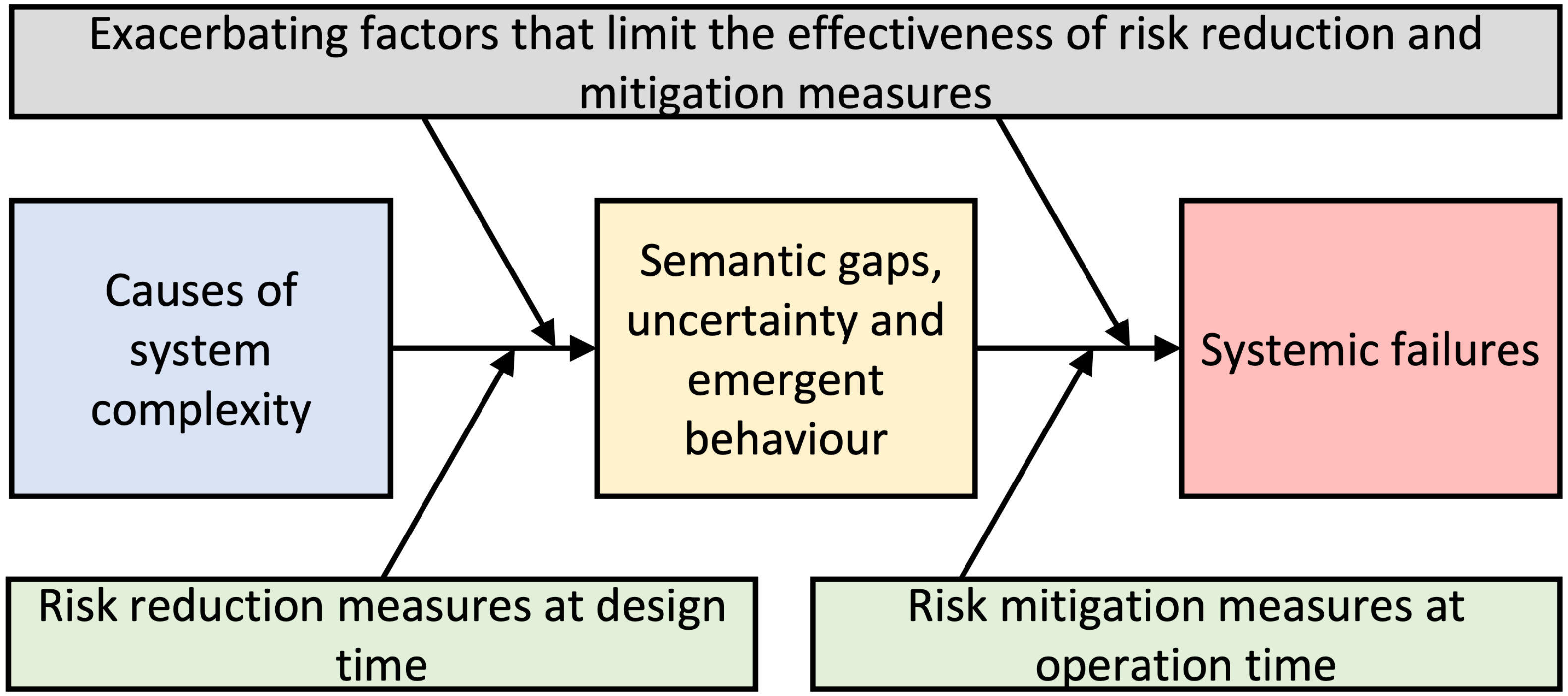
Lloyd's Register Foundation

EngineeringX

The Framework

Safety Causes, Consequences and Controls

- Framework recognises that failures (can) arise from complexity, rather than ”classical” failures
  - Exacerbating factors, akin to common cause failures
  - Need both operational and design time controls – not new, but …



Exacerbating factors that limit the effectiveness of risk reduction and mitigation measures

Causes of system complexity

Semantic gaps, uncertainty and emergent behaviour

Systemic failures

Risk reduction measures at design time

Risk mitigation measures at operation time

Use labels: causes, consequences and systemic failures

The Framework

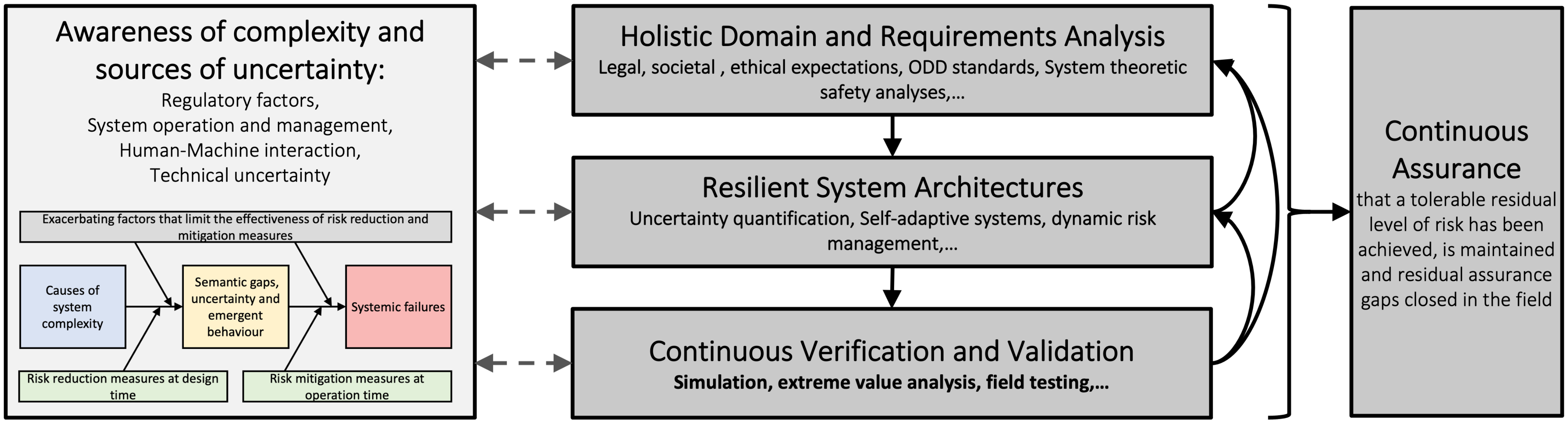
Layers

- Governance (& Regulation)
  - Cross-jurisdictional incentives and requirements for organisations to adhere to best practice through direct regulation, soft law approaches or a consensus in the form of national and international standards.
- Management
  - Risk management and informed design trade-offs including, management of supply chain dynamics and the sustainment of long-term institutional knowledge for long-lived and evolving systems.
- Technical & Human Factors (Task & Technical)
  - The technological components and the tasks performed by the users, operators and stakeholders within a socio-technical context.

The Framework

Towards Continuous Assurance

- The dynamics of complex systems, and the interaction between the layers, require a move towards continuous (continual) assurance



Awareness of complexity and sources of uncertainty:
Regulatory factors,
System operation and management,
Human-Machine interaction,
Technical uncertainty

Holistic Domain and Requirements Analysis
Legal, societal , ethical expectations, ODD standards, System theoretic safety analyses,...

Resilient System Architectures
Uncertainty quantification, Self-adaptive systems, dynamic risk management,...

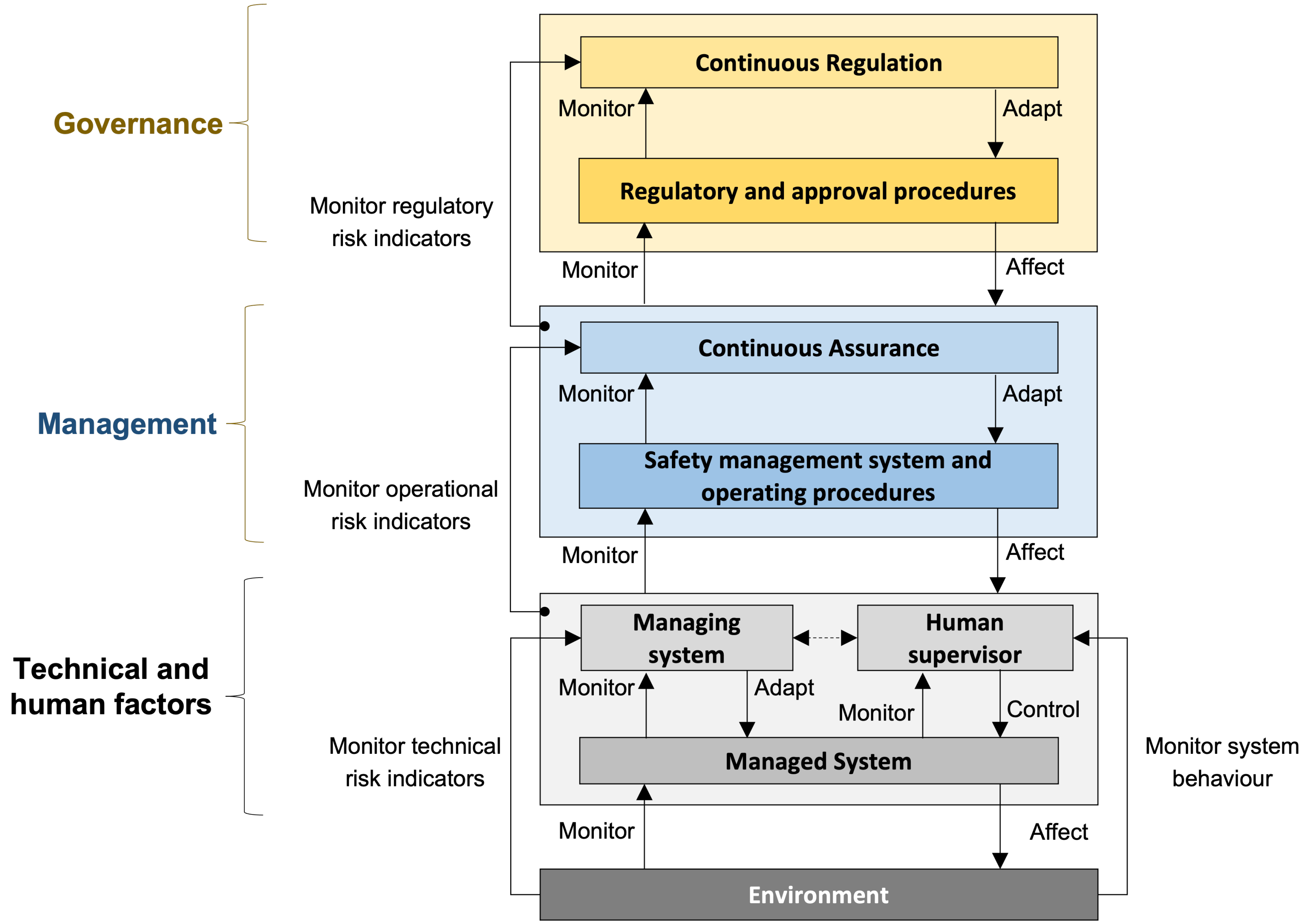
Continuous Verification and Validation
Simulation, extreme value analysis, field testing,...

Continuous Assurance
that a tolerable residual level of risk has been achieved, is maintained and residual assurance gaps closed in the field

The Framework

Towards Continuous Assurance

- Managing safety requires feedback across the layers
  - Potentially very rapid, in some cases
  - Further challenges including visibility in supply chain, and “pace” for regulators



Agenda

- Safer Complex Systems Study
- Defining Complex Systems
- Challenges
- A Framework for Managing Safety
- Some Examples
- Safety and Security
- Some Principles
- Conclusions

Engineering X

Safer Complex Systems

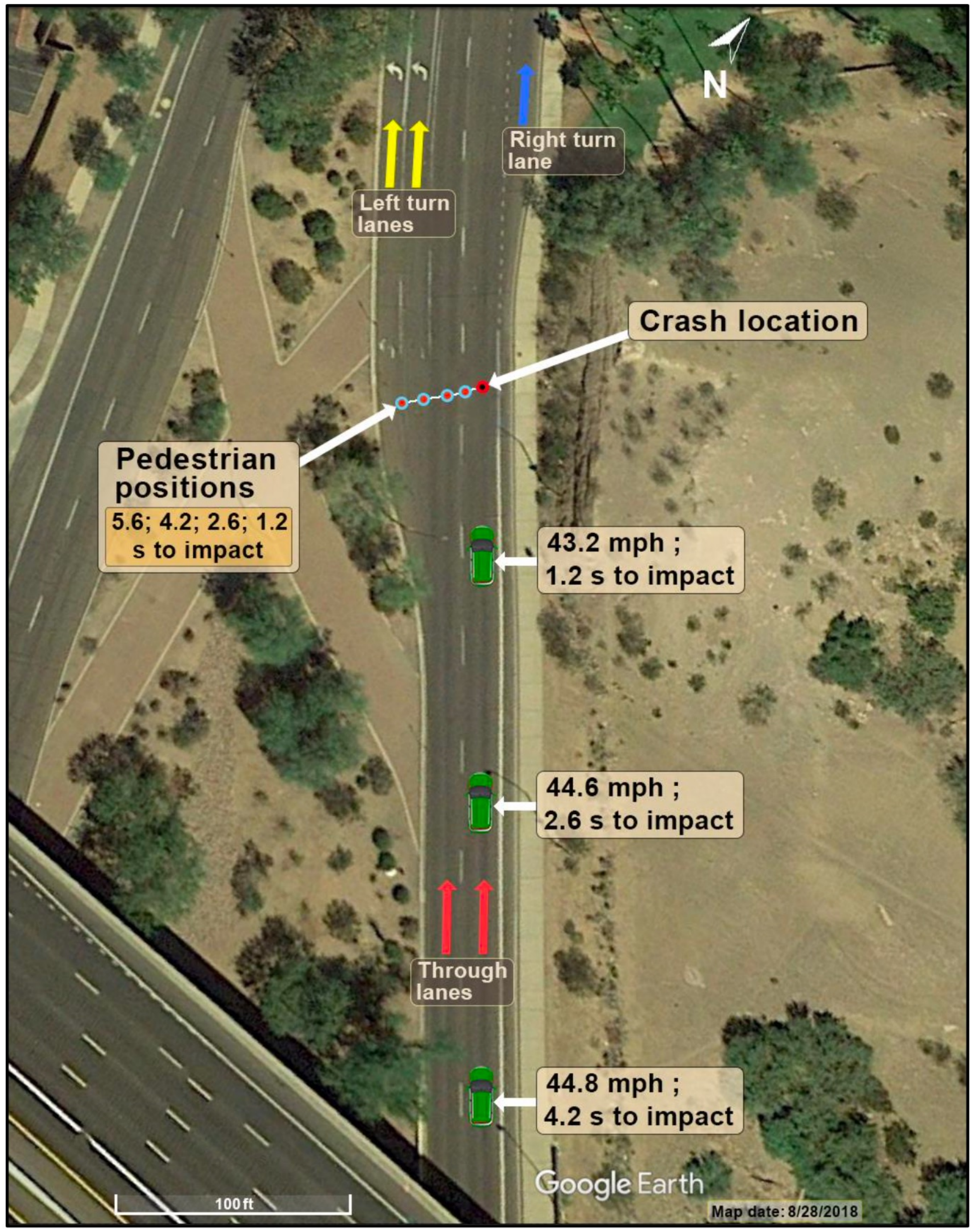
An Initial Framework

Royal Academy of Engineering

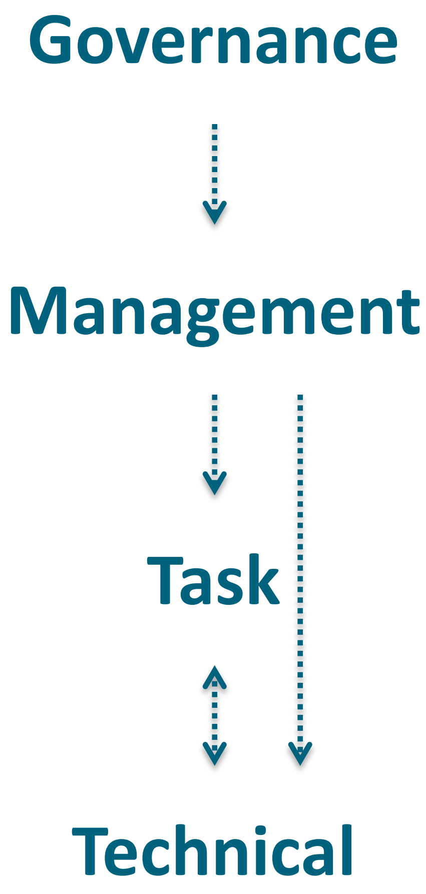
Lloyd's Register Foundation

Automotive Examples

Uber Tempe



Source: National Transportation Safety Board. Collision between vehicle controlled by developmental automated driving system and pedestrian Tempe, Arizona march 18, 2018. 2019.



Systemic Failures

Governance: Failure to regulate accountability for safety of automated driving

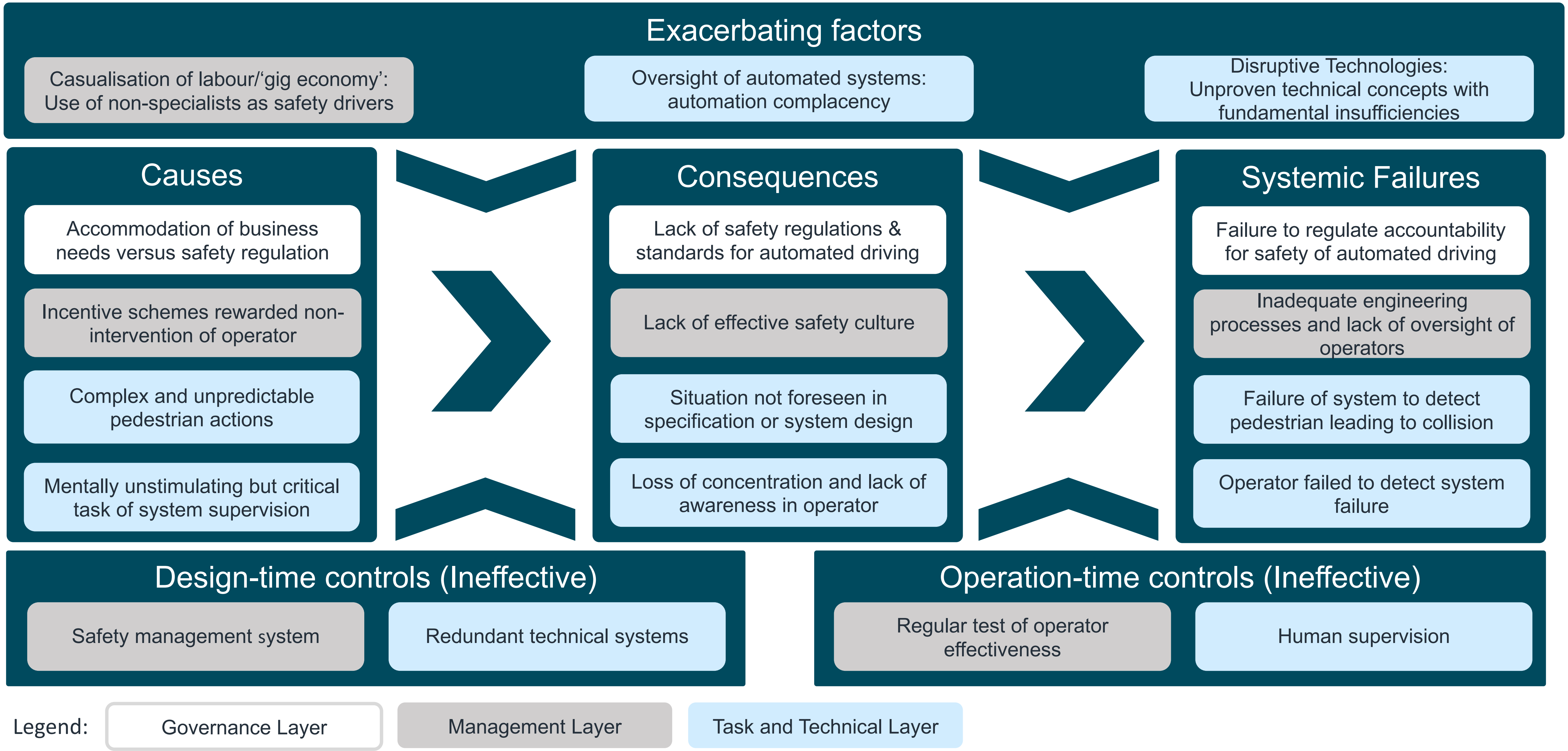
Management: Inadequate engineering processes and lack of oversight of operators

Task: Failure of operator to detect that system was not operating correctly

Technical: Failure of system to correctly detect pedestrian and avoid collision

Automotive Examples

Uber Tempe



Exacerbating factors

- Casualisation of labour/‘gig economy’: Use of non-specialists as safety drivers
- Oversight of automated systems: automation complacency
- Disruptive Technologies: Unproven technical concepts with fundamental insufficiencies

Causes

- Accommodation of business needs versus safety regulation
- Incentive schemes rewarded non-intervention of operator
- Complex and unpredictable pedestrian actions
- Mentally unstimulating but critical task of system supervision

Consequences

- Lack of safety regulations & standards for automated driving
- Lack of effective safety culture
- Situation not foreseen in specification or system design
- Loss of concentration and lack of awareness in operator

Systemic Failures

- Failure to regulate accountability for safety of automated driving
- Inadequate engineering processes and lack of oversight of operators
- Failure of system to detect pedestrian leading to collision
- Operator failed to detect system failure

Design-time controls (Ineffective)

- Safety management system
- Redundant technical systems

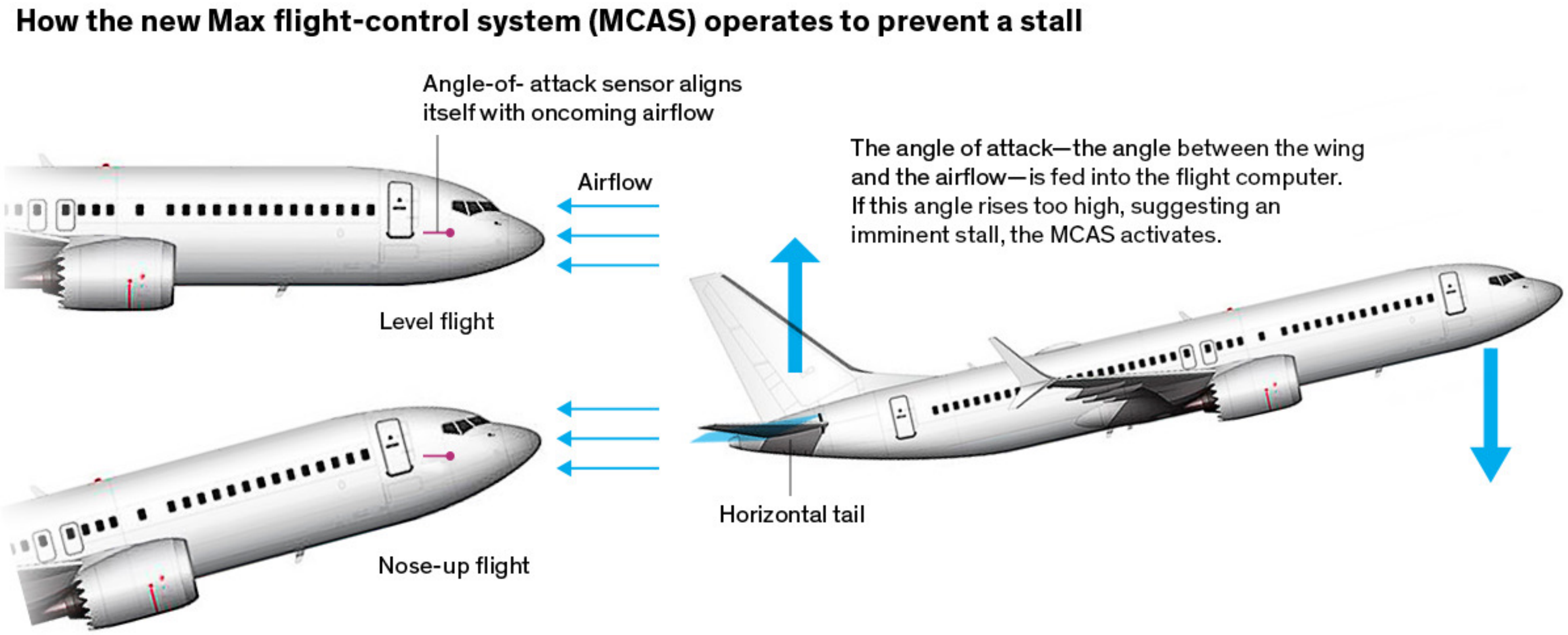
Operation-time controls (Ineffective)

- Regular test of operator effectiveness
- Human supervision

Legend: Governance Layer | Management Layer | Task and Technical Layer

Aerospace Examples

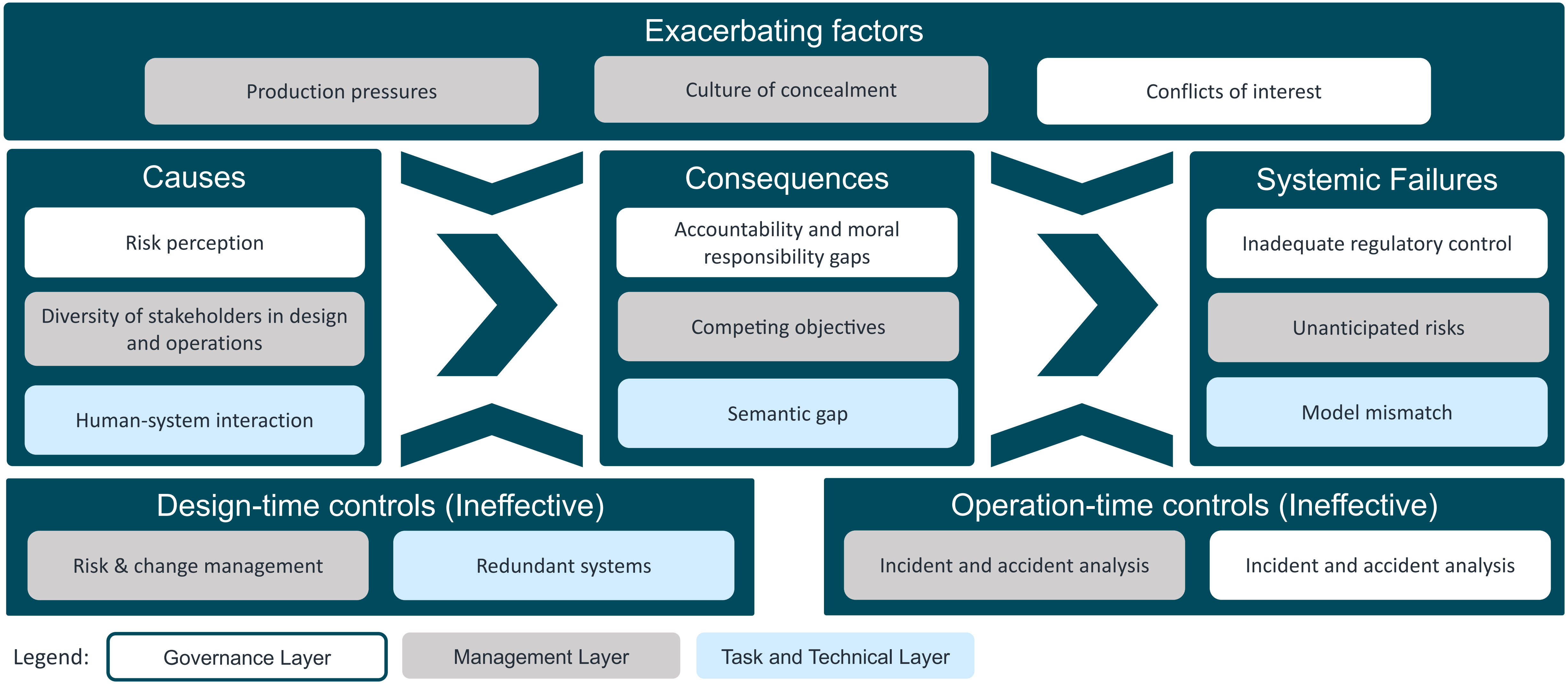
737 Max



- Watch congressional hearings – not just technical

Aerospace Examples

737 Max



Exacerbating factors

- Production pressures
- Culture of concealment
- Conflicts of interest

Causes

- Risk perception
- Diversity of stakeholders in design and operations
- Human-system interaction

Consequences

- Accountability and moral responsibility gaps
- Competing objectives
- Semantic gap

Systemic Failures

- Inadequate regulatory control
- Unanticipated risks
- Model mismatch

Design-time controls (Ineffective)

- Risk & change management
- Redundant systems

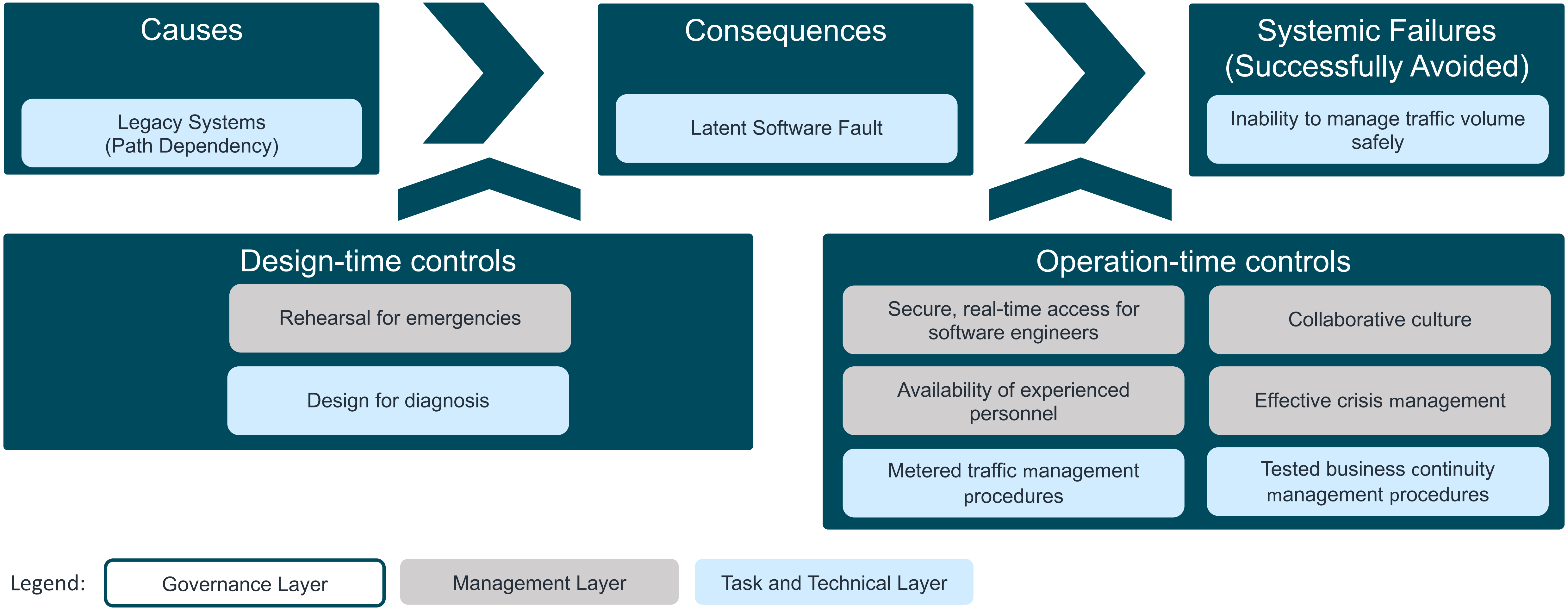
Operation-time controls (Ineffective)

- Incident and accident analysis
- Incident and accident analysis

Legend: Governance Layer | Management Layer | Task and Technical Layer

Aerospace Examples

NATS Outage



Causes

Legacy Systems
(Path Dependency)

Consequences

Latent Software Fault

Systemic Failures
(Successfully Avoided)

Inability to manage traffic volume safely

Design-time controls

Rehearsal for emergencies

Design for diagnosis

Operation-time controls

Secure, real-time access for software engineers

Collaborative culture

Availability of experienced personnel

Effective crisis management

Metered traffic management procedures

Tested business continuity management procedures

Legend: Governance Layer | Management Layer | Task and Technical Layer

- Example of successful management
  - Mainly operational controls

UNIVERSITY of York

Agenda

- Safer Complex Systems Study
- Defining Complex Systems
- Challenges
- A Framework for Managing Safety
- Some Examples
- Safety and Security
- Some Principles
- Conclusions

Engineering X

Safer Complex Systems

An Initial Framework

Royal Academy of Engineering

Lloyd's Register Foundation

Engineering X

Security and DevOps

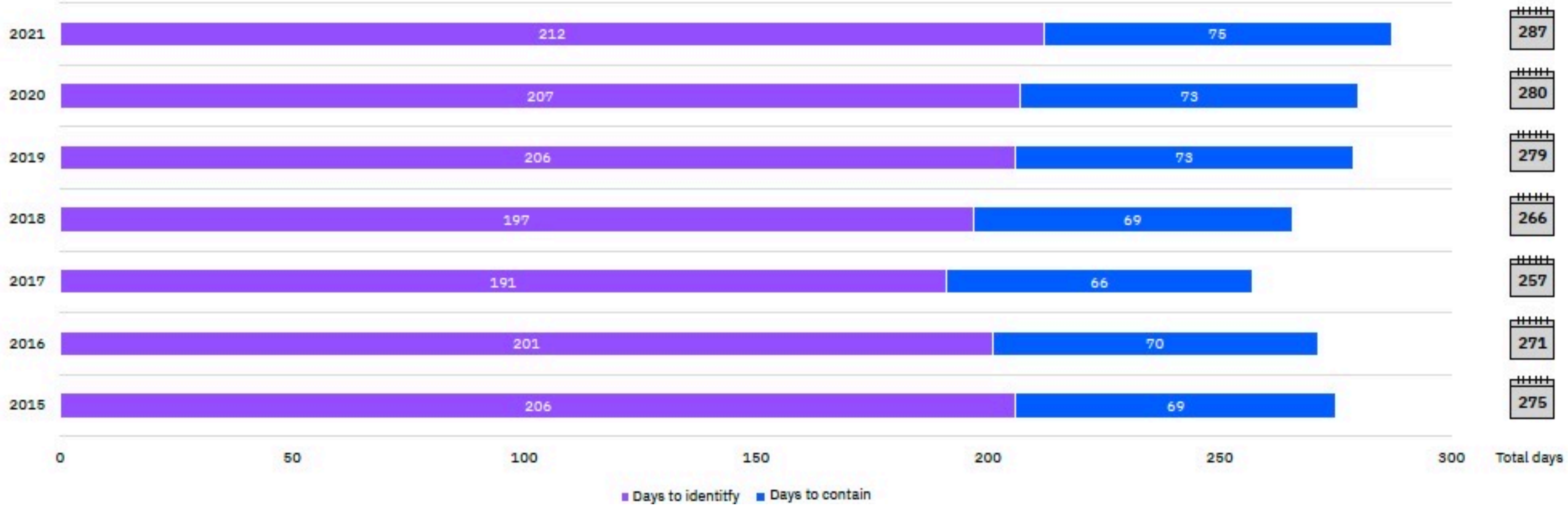
Identifying and Repairing Breaches

- Data on DevOps not very encouraging
  - Despite the dynamism of DevOps, very long response times

Figure 9

Average time to identify and contain a data breach

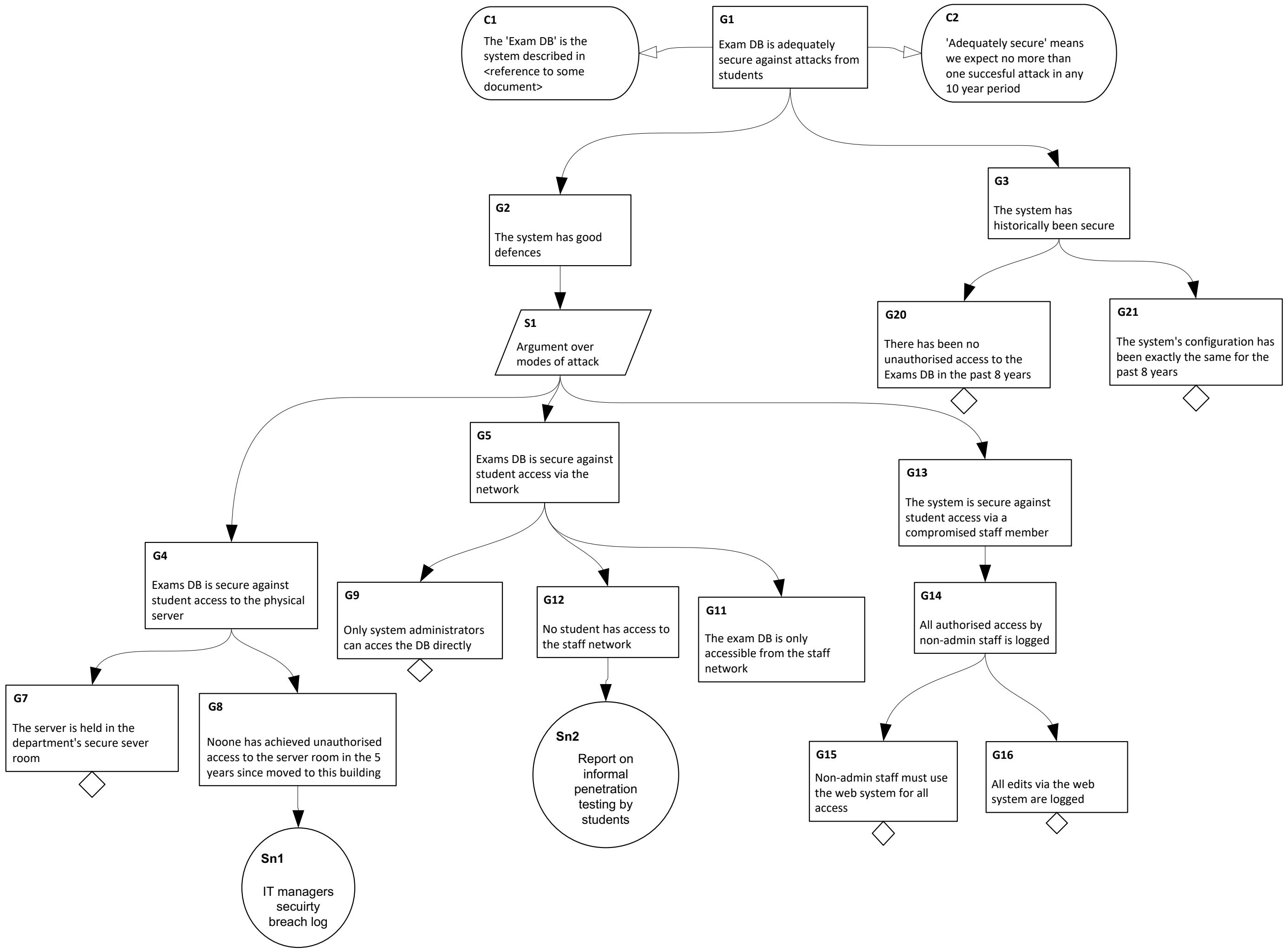
Measured in days



Safety and Security

Analysis and Argument

- Need to take an integrated view of safety and security
  - Analysis methods so security breaches are considered as potential hazard causes
    - Early life-cycle to drive design, & confirmation near the end
  - Safety and security trade-offs
  - Safety and security assurance cases
  - Need to consider dynamics …



C1
The 'Exam DB' is the system described in <reference to some document>

G1
Exam DB is adequately secure against attacks from students

C2
'Adequately secure' means we expect no more than one succesful attack in any 10 year period

G2
The system has good defences

G3
The system has historically been secure

S1
Argument over modes of attack

G20
There has been no unauthorised access to the Exams DB in the past 8 years

G21
The system's configuration has been exactly the same for the past 8 years

G5
Exams DB is secure against student access via the network

G13
The system is secure against student access via a compromised staff member

G4
Exams DB is secure against student access to the physical server

G9
Only system administrators can acces the DB directly

G12
No student has access to the staff network

G11
The exam DB is only accessible from the staff network

G14
All authorised access by non-admin staff is logged

G7
The server is held in the department's secure sever room

G8
Noone has achieved unauthorised access to the server room in the 5 years since moved to this building

Sn2
Report on informal penetration testing by students

G15
Non-admin staff must use the web system for all access

G16
All edits via the web system are logged

Sn1
IT managers secuirty breach log

Agenda

- Safer Complex Systems Study
- Defining Complex Systems
- Challenges
- A Framework for Managing Safety
- Some Examples
- Safety and Security
- Some Principles
- Conclusions

Engineering X

Safer Complex Systems

An Initial Framework

Royal Academy of Engineering

Lloyd's Register Foundation

Some Principles

De Minimis?

- Resilience
  - Need to design for observability (NATS, DevOps), but NB AI
  - Need to design for human controls
  - Need to rehearse, but NB ad hoc systems
- Dynamics
  - Monitor systems to identify leading and lagging indicators
  - Need to update safety and security assessments dynamically
  - Prompt action if needed, at appropriate layer
- Take a managerial view
  - Consider from the point of view of a safety/security management, "drive out" requirements for operational controls (to influence design)

Agenda

- Safer Complex Systems Study
- Defining Complex Systems
- Challenges
- A Framework for Managing Safety
- Some Examples
- Safety and Security
- Some Principles
- Conclusions

EngineeringX

Safer Complex Systems

An Initial Framework

Royal Academy of Engineering

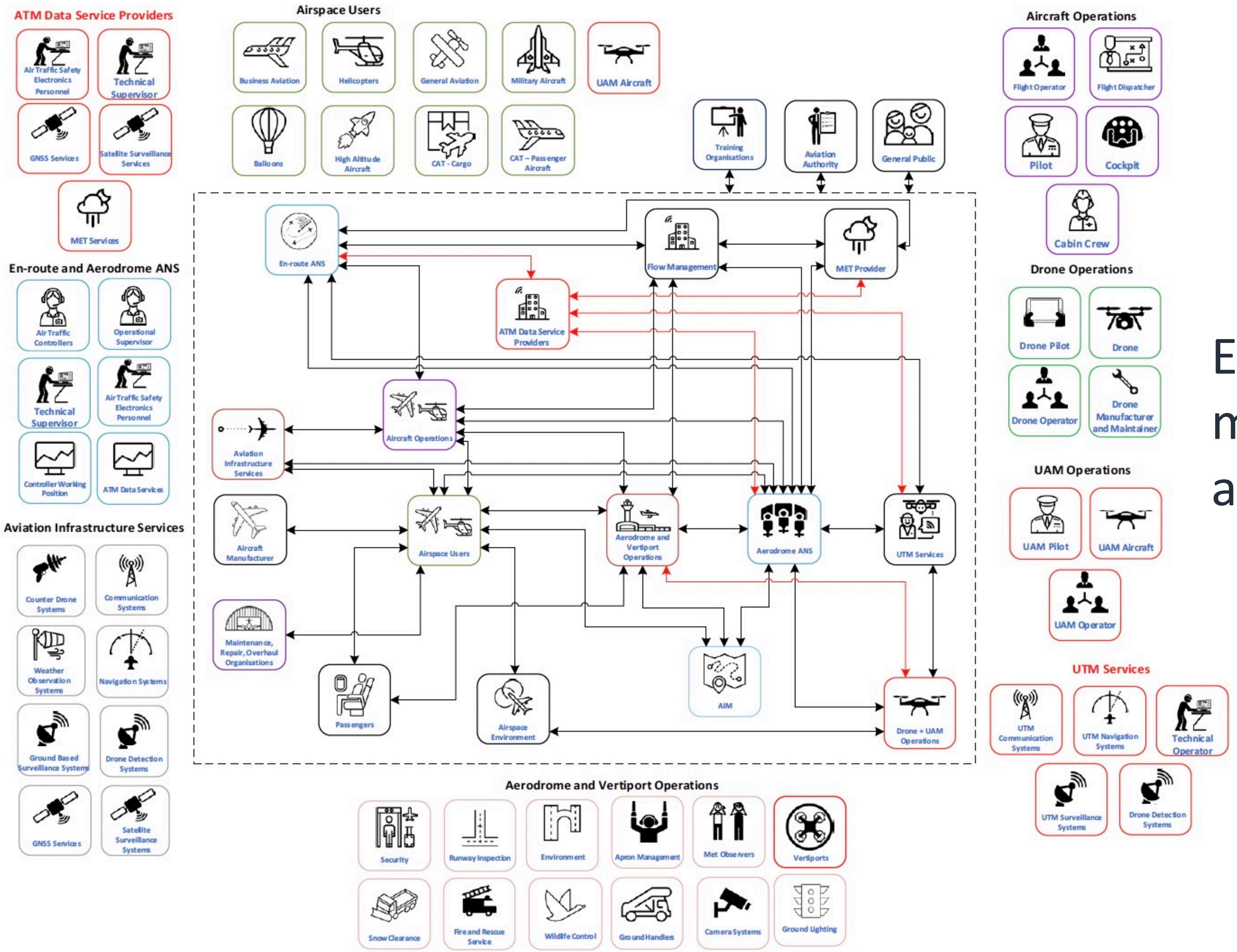
Lloyd's Register Foundation

Conclusions

Managing Complexity

- Framework from the Safer Complex Systems programme
  - Largely descriptive, but helps by providing a holistic point of view
  - A few examples of use, mainly post hoc or to describe situations
- More work needed
  - Refinement of the framework and guidance in its use
  - Examples that can help to drive design
  - Analysis “methods”, e.g. extending STAMP/STPA with additional prompts, and combined safety-security analyses
  - Effective ways of putting (de minimis) principles into practice
  - Approaches to system (system of systems) modelling to aid approach

Example SoS Model: ATM



Evolution towards more autonomous air services

https://www.egis-group.com/projects/future-flight

The Project Team

Professor John McDermid OBE FREng

- Director, Assuring Autonomy International Programme
- University of York

Professor Simon Burton

- Project Director, Fraunhofer IKS, Munich
- Ex-Director Vehicle Systems Safety, Robert Bosch GmbH
- Visiting Professor University of York

Dr Philip Garnett

- Senior Lecturer in Systems and Organisation, member of YCCSA and Co-director of SATSU
- University of York

Dr Rob Weaver

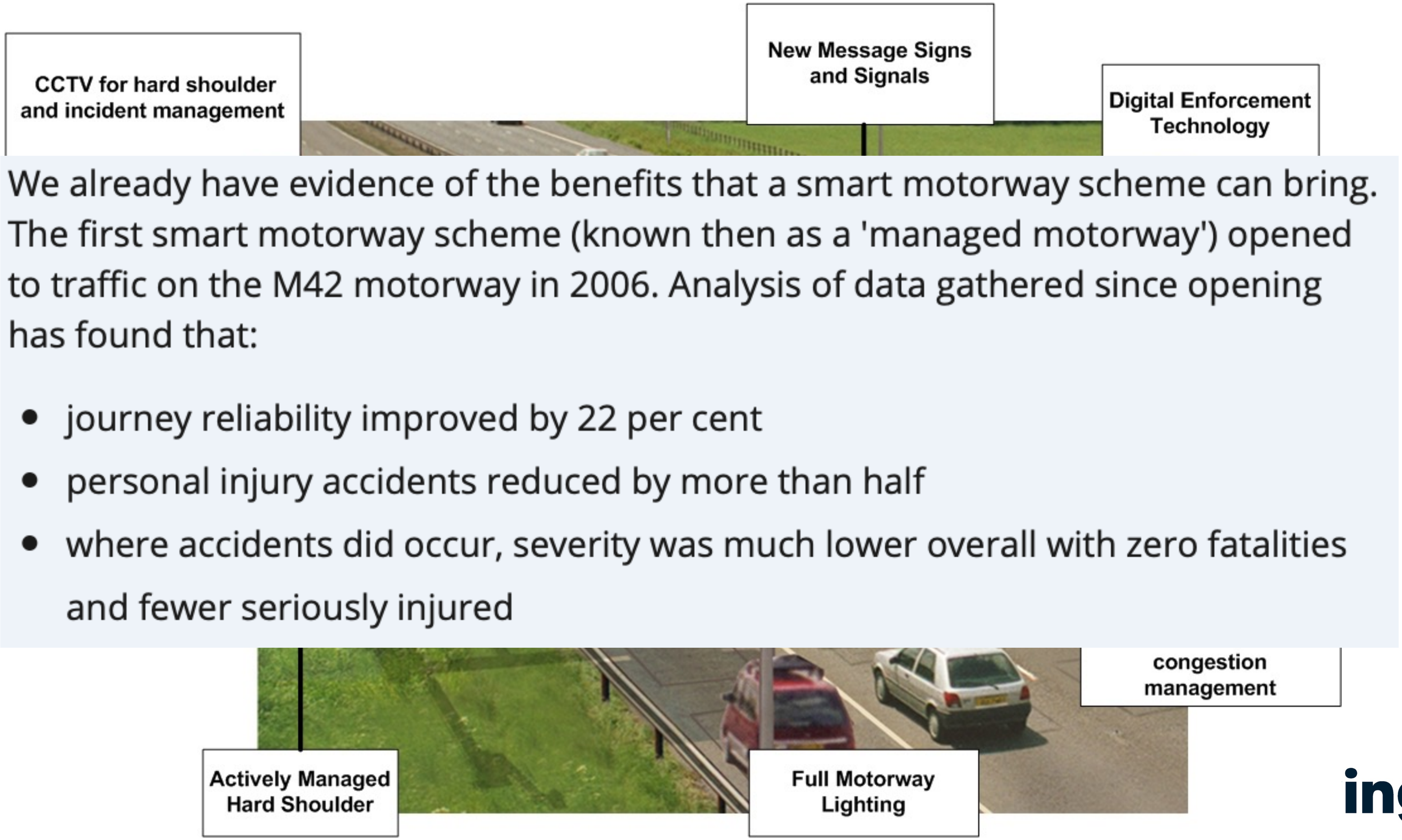
- Global Aviation and Safety Advisor, working on future traffic management concepts and urban air mobility
- Former Head of Safety for Australian Air Traffic Control

Questions and Discussion

Contact: john.mcdermid@york.ac.uk

Some Examples

Smart Motorways – M42

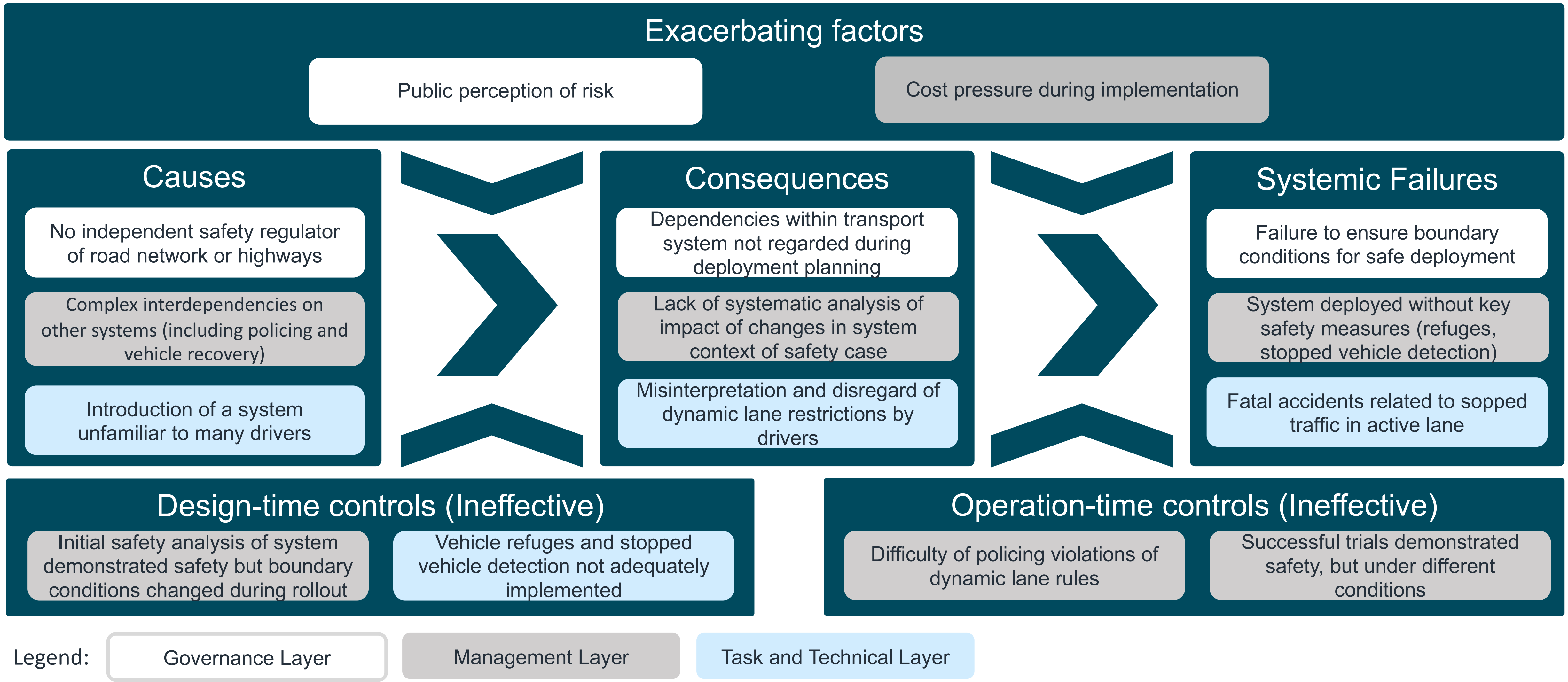


We already have evidence of the benefits that a smart motorway scheme can bring. The first smart motorway scheme (known then as a 'managed motorway') opened to traffic on the M42 motorway in 2006. Analysis of data gathered since opening has found that:

- journey reliability improved by 22 per cent
- personal injury accidents reduced by more than half
- where accidents did occur, severity was much lower overall with zero fatalities and fewer seriously injured

Some Examples

Smart Motorways – Rollout



Exacerbating factors

Public perception of risk

Cost pressure during implementation

Causes

No independent safety regulator of road network or highways

Complex interdependencies on other systems (including policing and vehicle recovery)

Introduction of a system unfamiliar to many drivers

Consequences

Dependencies within transport system not regarded during deployment planning

Lack of systematic analysis of impact of changes in system context of safety case

Misinterpretation and disregard of dynamic lane restrictions by drivers

Systemic Failures

Failure to ensure boundary conditions for safe deployment

System deployed without key safety measures (refuges, stopped vehicle detection)

Fatal accidents related to sopped traffic in active lane

Design-time controls (Ineffective)

Initial safety analysis of system demonstrated safety but boundary conditions changed during rollout

Vehicle refuges and stopped vehicle detection not adequately implemented

Operation-time controls (Ineffective)

Difficulty of policing violations of dynamic lane rules

Successful trials demonstrated safety, but under different conditions

Legend: Governance Layer | Management Layer | Task and Technical Layer