ASSURING
AUTONOMY
INTERNATIONAL PROGRAMME

Dynamic Reasoning for Safety Assurance

Ibrahim Habli
Ibrahim.habli@york.ac.uk

Based on an ICSE NIER 2015 paper with Ewen Denney and Ganesh Pai
https://ti.arc.nasa.gov/publications/21593/download

Background

- Paradigm shift in many domains
  - Shift from a prescribed process to a product-oriented assurance
  - Shift from a tick-box to argument-based

- Different drivers:
  - Accidents
    - Piper Alpha, 1988
  - Different business model
    - Rail privatisation, 1992
  - Incidents and recalls
    - FDA, 2010
  - Complexity
    - Automotive, 2011

Safety Case Contents

Safety argument typically depends on:

1. Specification of a particular **system design**
2. Description of a particular **configuration** and **environment** in which the design will operate
3. An identified list of **hazards** associated with system operation
4. A claim that the list of hazards is **sufficient**
5. An assessment of the safety **risk** presented by each hazard, including *estimates* and *assumptions* used for quantification
6. Claims about the **effectiveness** of the chosen risk **mitigation** measures
7. A claim that for all mitigations which were not included, the mitigations were not reasonably **practicable** to implement

[Rae 2009]

All of the above can, and often, change

Translating Tensions into Safe Practices Through Dynamic Trade-offs: The Secret Second Handover

Mark A Sujan, Peter Spurgeon and Matthew W Cooke

Introduction

Failures in the handover of responsibility for patient care from one caregiver to another are a recognised threat to patient safety (Cohen and Hilligoss, 2010; Raduma-Tomas, Flin, Yule and Williams, 2011). Handover in emergency care comes in different shapes and forms. For example, at the end of a shift there is a handover between the outgoing physician and the incoming party. This type of handover takes place between individuals of the same discipline and professional background. In addition, there are other types of handover that take place along the patient pathway. For example, ambulance crews hand over to emergency department (ED) staff. At the other end of the patient pathway in emergency care there are hand-overs from ED staff to inpatient hospital services. The study presented in this chapter focused on these types of handover along the patient pathway, because they involve individuals from different departments and organisations who have different professional, organisational and cultural backgrounds (Sujan et al., 2013). This diversity makes handover across care boundaries even more challenging and error prone than shift hand-overs (Hilligoss and Cohen, 2013).

A key finding of the study of handover in emergency care is that performance variability of front-line practitioners is an essential component in the delivery of safe care (Sujan et al., 2014). Performance variability in the context of handover does not suggest that handover practices should be unreliable, random or haphazard. Rather, the research found that practitioners adapt their behaviour and their practices based on their experience and expertise, and depending on the characteristics of the specific situation. The resulting behaviours are an example of how work-as-done (WAD) by practitioners is necessarily different from work-as-imagined (WAI) by system designers and managers (Hollnagel, Braithwaite and Wears, 2013). These adaptations by practitioners are important because there are tensions or contradictions in patient handover and the care processes more generally. Practitioners translate these tensions or contradictions into safe practices through dynamic trade-offs on a daily basis as part of their everyday clinical work. We propose that this process creates, or is a mechanism of, resilience.

In this chapter, we explore such tensions and dynamic trade-offs through an example from our research on the safety of handover across care boundaries in emergency care. The next section describes the case study. We then discuss the key theoretical concepts and their relationship to Resilience Engineering. We conclude the chapter with implications for research and for practice.

Case Study: Handover in Emergency Care

In this section we provide from our fieldwork one illustrative example of how practitioners make dynamic

Resilience or Safety 2.0

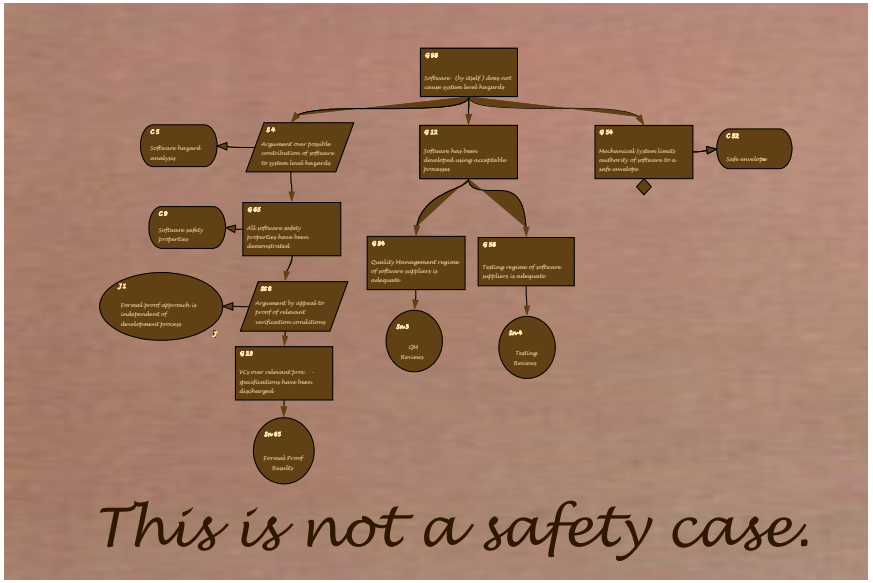
The intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances, so that it can sustain required operations under both expected and unexpected conditions.

Erik Hollnagel

Safety Case Depictions vs. Safety Case Reports

Would the Real Safety Case Please Stand Up?

Ibrahim Habli, Tim Kelly, 2007

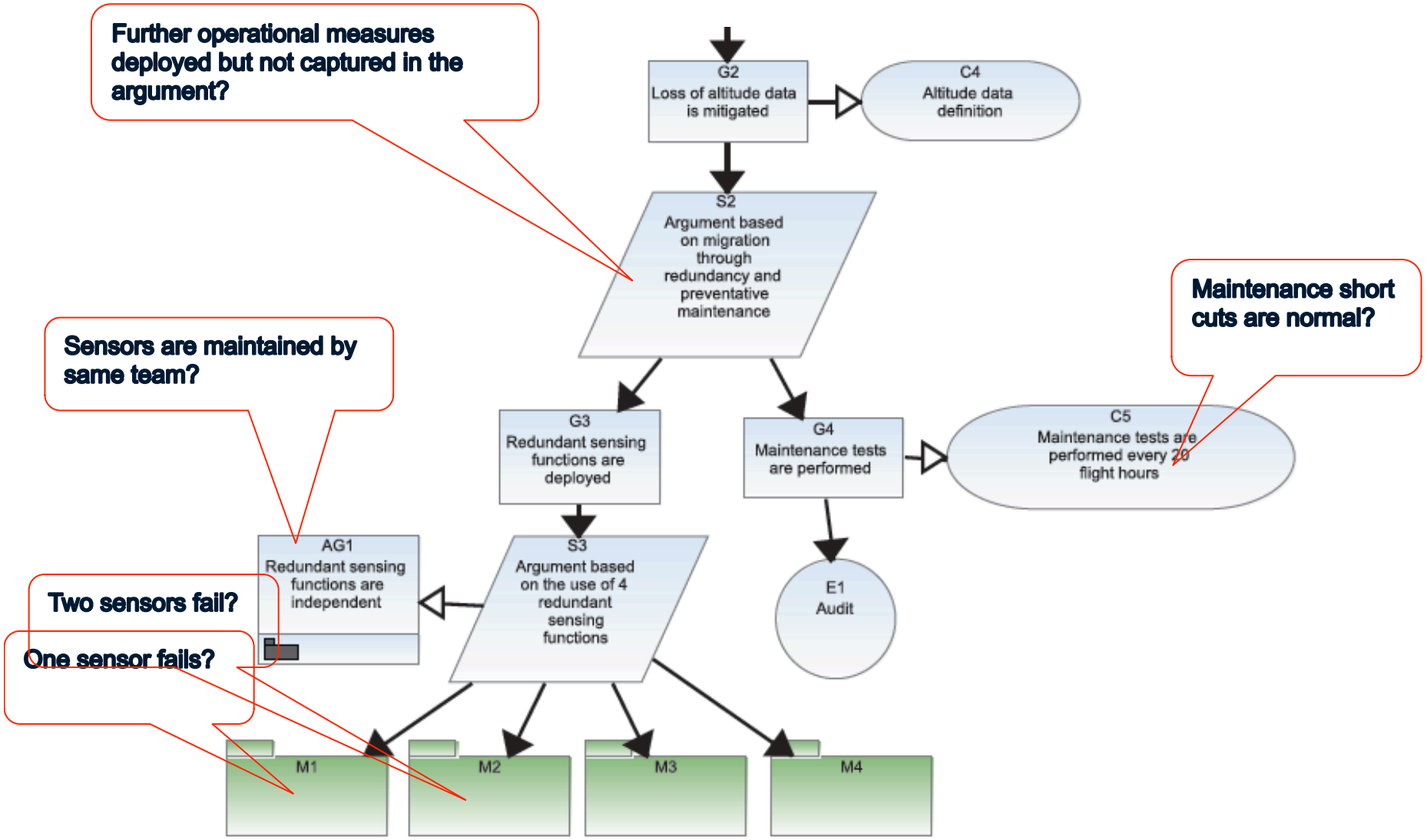


Difference between the *actual* and the *depicted*

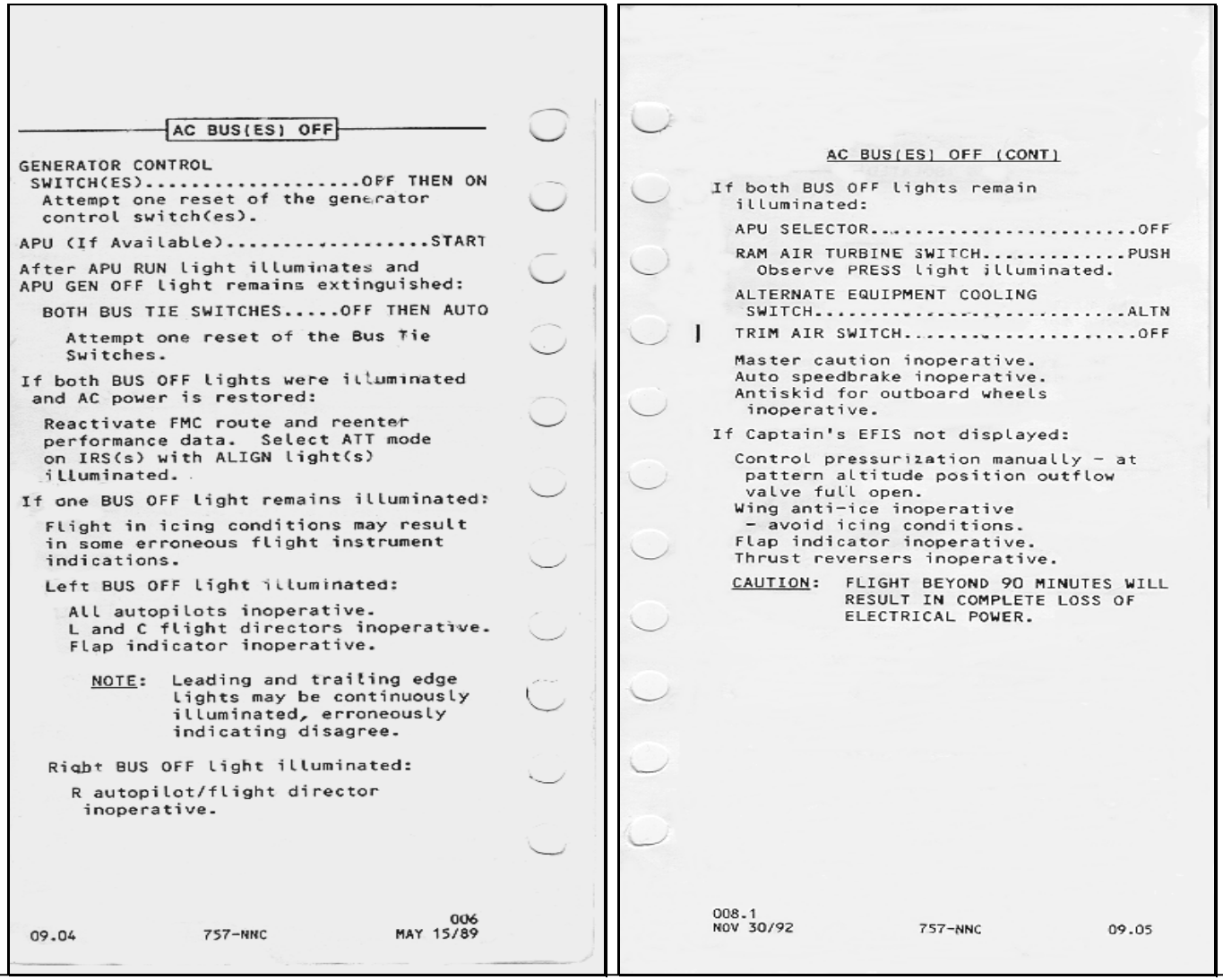
The gap can lead to "**a culture of 'paper safety' at the expense of real safety**".

[Inquiry Report following the RAF Nimrod aircraft accident]

Example



QRH pages from Boeing B-757



AC BUS(ES) OFF

GENERATOR CONTROL SWITCH(ES).....OFF THEN ON
Attempt one reset of the generator control switch(es).

APU (If Available).....START

After APU RUN light illuminates and APU GEN OFF light remains extinguished:

BOTH BUS TIE SWITCHES.....OFF THEN AUTO

Attempt one reset of the Bus Tie Switches.

If both BUS OFF lights were illuminated and AC power is restored:

Reactivate FMC route and reenter performance data. Select ATT mode on IRS(s) with ALIGN light(s) illuminated.

If one BUS OFF light remains illuminated:

Flight in icing conditions may result in some erroneous flight instrument indications.

Left BUS OFF light illuminated:

All autopilots inoperative.
L and C flight directors inoperative.
Flap indicator inoperative.

NOTE: Leading and trailing edge lights may be continuously illuminated, erroneously indicating disagree.

Right BUS OFF light illuminated:

R autopilot/flight director inoperative.

AC BUS(ES) OFF (CONT)

If both BUS OFF lights remain illuminated:

APU SELECTOR.....OFF

RAM AIR TURBINE SWITCH.....PUSH
Observe PRESS light illuminated.

ALTERNATE EQUIPMENT COOLING SWITCH.....ALTN

TRIM AIR SWITCH.....OFF

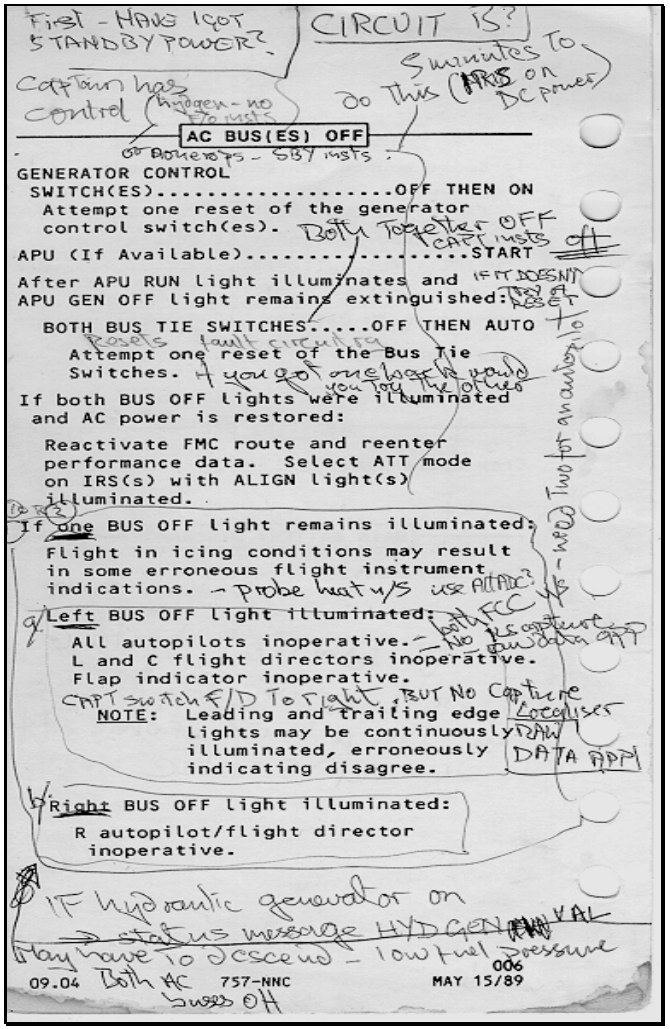
Master caution inoperative.
Auto speedbrake inoperative.
Antiskid for outboard wheels inoperative.

If Captain's EFIS not displayed:

Control pressurization manually - at pattern altitude position outflow valve full open.
Wing anti-ice inoperative - avoid icing conditions.
Flap indicator inoperative.
Thrust reversers inoperative.

CAUTION: FLIGHT BEYOND 90 MINUTES WILL RESULT IN COMPLETE LOSS OF ELECTRICAL POWER.

Same QRH pages WITH pilot annotations



AC BUS(ES) OFF

GENERATOR CONTROL SWITCH(ES).....OFF THEN ON
Attempt one reset of the generator control switch(es).

APU (If Available).....START

After APU RUN light illuminates and APU GEN OFF light remains extinguished:

BOTH BUS TIE SWITCHES.....OFF THEN AUTO
Attempt one reset of the Bus Tie Switches.

If both BUS OFF lights were illuminated and AC power is restored:

Reactivate FMC route and reenter performance data. Select ATT mode on IRS(s) with ALIGN light(s) illuminated.

If one BUS OFF light remains illuminated:

Flight in icing conditions may result in some erroneous flight instrument indications.

Left BUS OFF light illuminated:

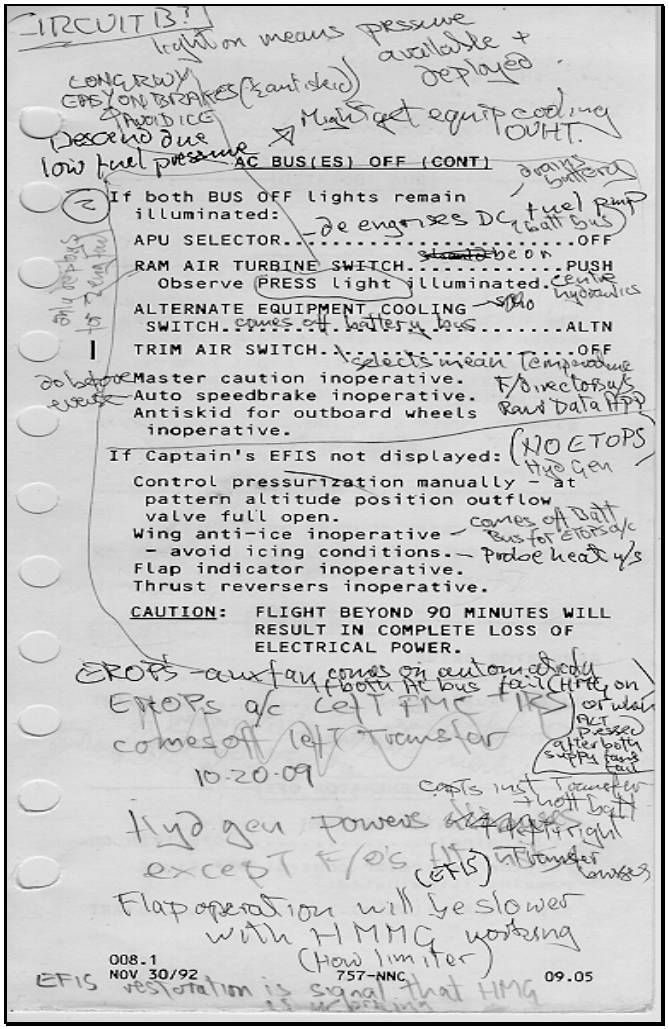
All autopilots inoperative.
L and C flight directors inoperative.
Flap indicator inoperative.

NOTE: Leading and trailing edge lights may be continuously illuminated, erroneously indicating disagree.

Right BUS OFF light illuminated:

R autopilot/flight director inoperative.

09.04 757-NNC MAY 15/89 006



AC BUS(ES) OFF (CONT)

If both BUS OFF lights remain illuminated:

APU SELECTOR.....OFF

RAM AIR TURBINE SWITCH.....PUSH
Observe PRESS light illuminated.

ALTERNATE EQUIPMENT COOLING SWITCH.....ALTN

TRIM AIR SWITCH.....OFF

Master caution inoperative.
Auto speedbrake inoperative.
Antiskid for outboard wheels inoperative.

If Captain's EFIS not displayed:

Control pressurization manually - at pattern altitude position outflow valve full open.
Wing anti-ice inoperative - avoid icing conditions.
Flap indicator inoperative.
Thrust reversers inoperative.

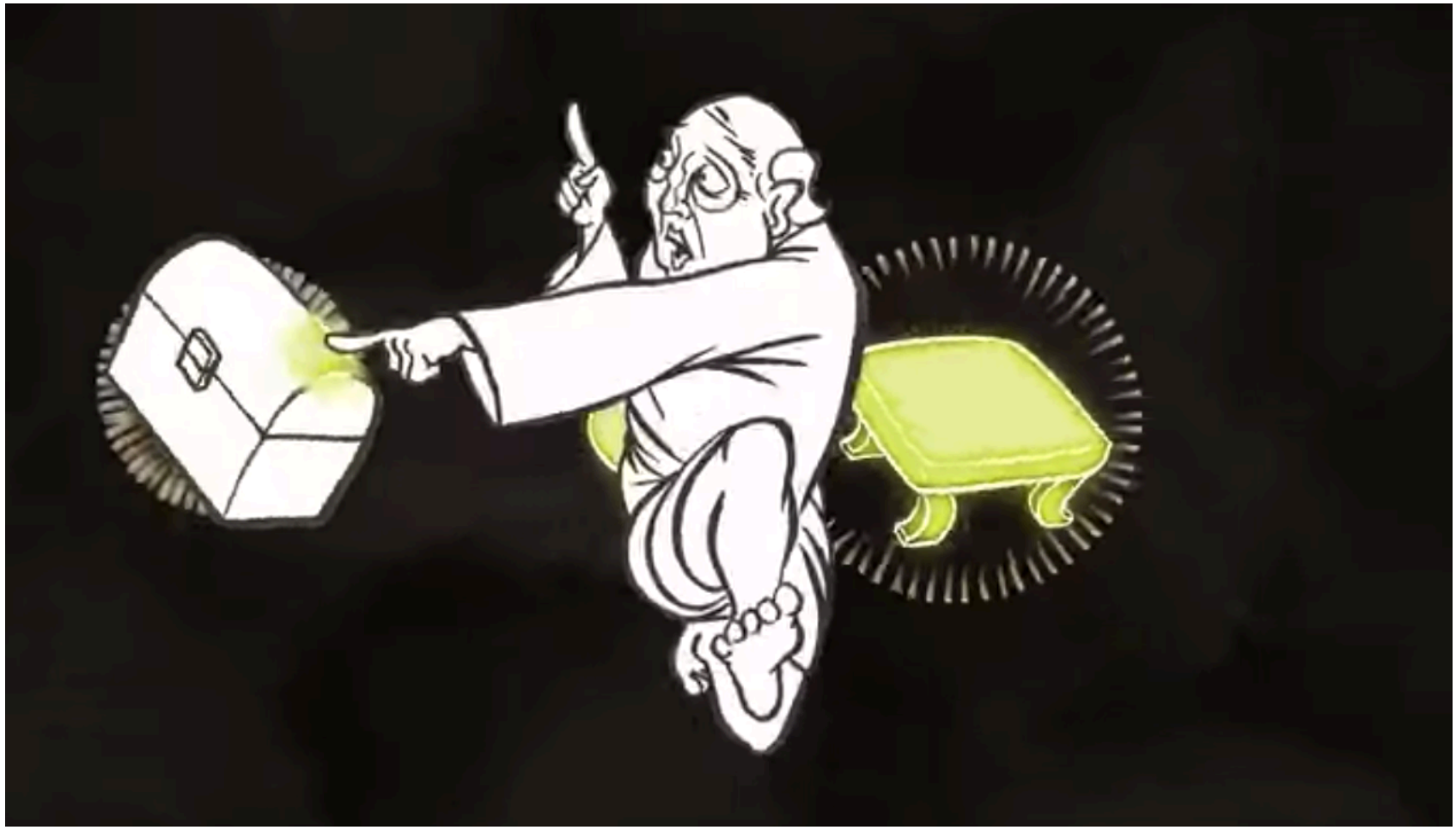
CAUTION: FLIGHT BEYOND 90 MINUTES WILL RESULT IN COMPLETE LOSS OF ELECTRICAL POWER.

008.1 NOV 30/92 757-NNC 09.05

Why is this important particularly now?

- Change in landscape of safety-critical applications
  - Increasing authority, autonomy, adaptation, and communication
  - Greater uncertainty about safe operation
    - including for historically stable domains such as aerospace and automotive

The Myth of King Midas and his Golden Touch



AI and Safety Requirements 1/2

- How do you specify cleanliness or making a cup of tea for a domestic robot?

[Building safe artificial intelligence: specification, robustness, and assurance by DeepMind]

AI and Safety Requirements 2/2

- Ideal requirements (“wishes”), corresponding to the hypothetical (but hard to articulate) description of an ideal AI system

- System/software requirements (“blueprint”), corresponding to the requirements that we actually use to build the AI system, e.g. a reward function to maximise

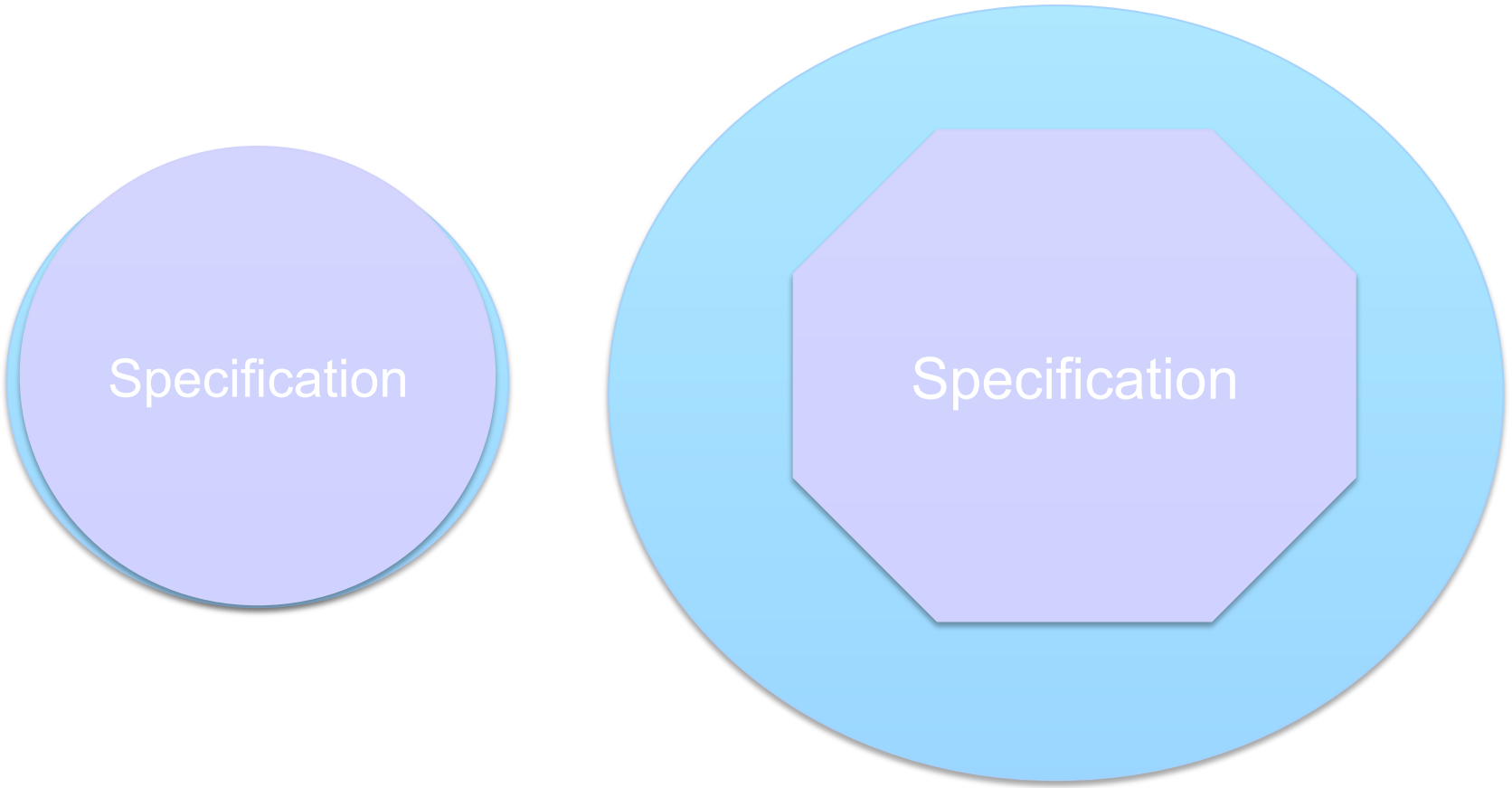
- Revealed requirements (“behaviour”), that best describes what actually happens, e.g. the reward function we can reverse-engineer from observing the system’s behaviour

How do we reduce the gap between the above?

[Building safe artificial intelligence: specification, robustness, and assurance by DeepMind]

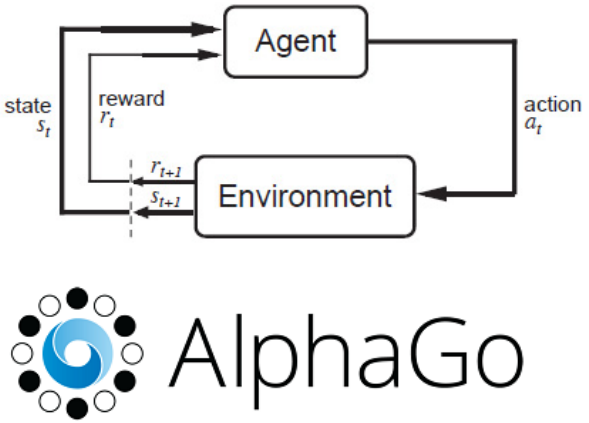
Autonomy and Intelligence

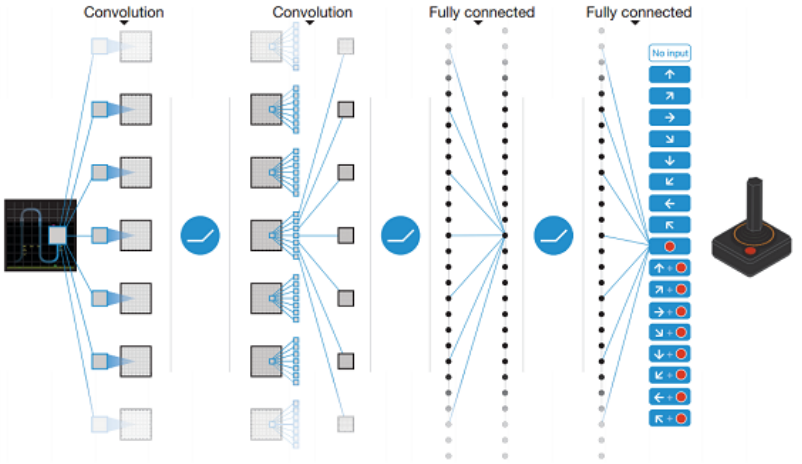
- Problem Domain



Autonomy and Intelligence

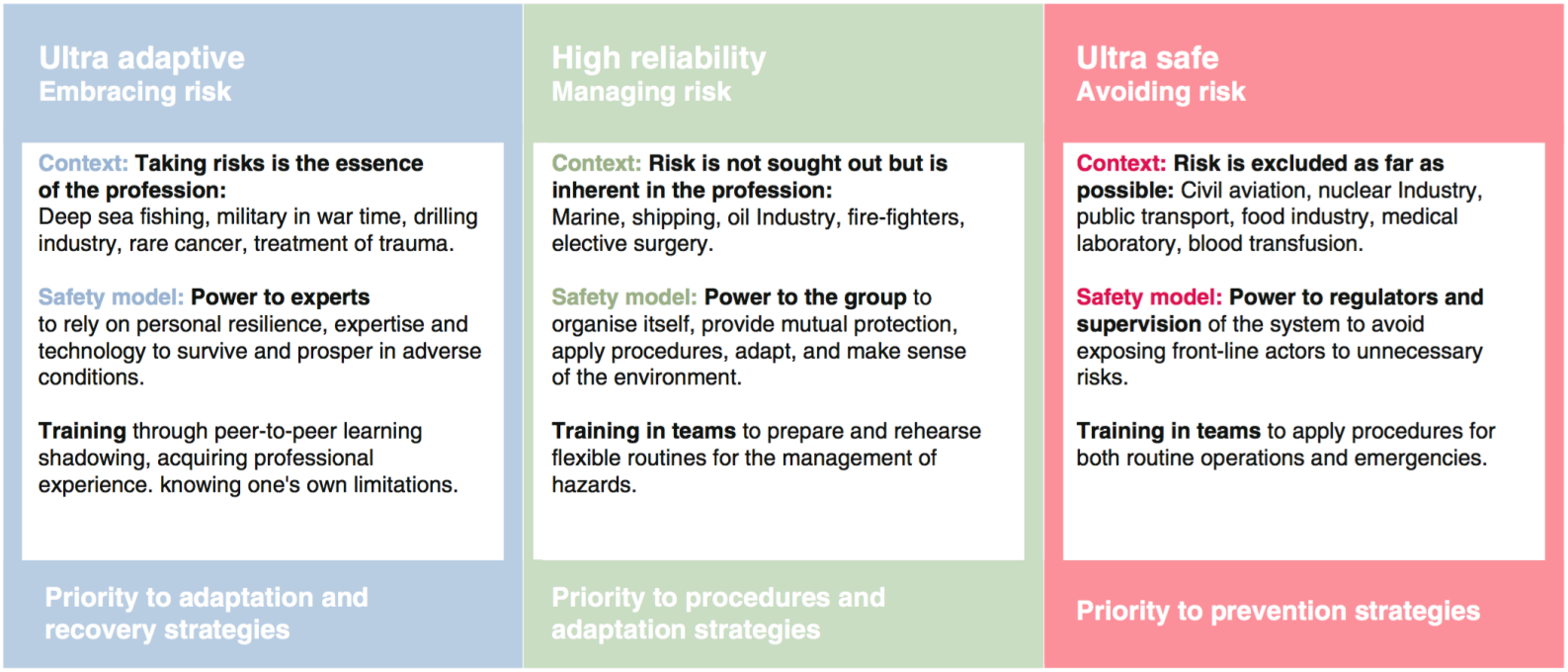
- Solution Domain





https://adeshpande3.github.io/Deep-Learning-Research-Review-Week-2-Reinforcement-Learning

Contrasting Approaches to Safety



[Vincent and Amalberti 2016]

Dynamic Safety Cases

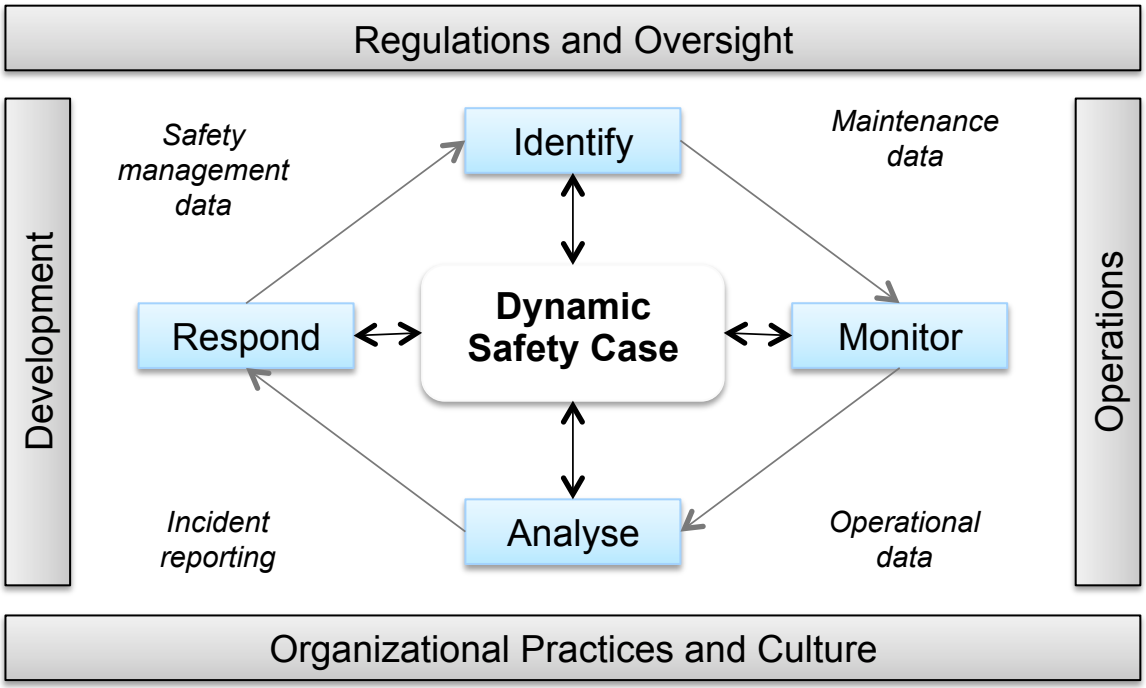
Aim of Dynamic Safety Cases

To **continuously compute** confidence in, and **proactively update** the reasoning about, the safety of ongoing operations

Attributes of Dynamic Safety Cases

- Continuity
  - safety is an operational concept
- Proactivity
  - deal with leading indicators of, and precursors to, hazardous behaviour
    - i.e not just faults and failures
- Computability
  - assessment of current confidence based on operational data
  - a high degree of automation and formality is necessary?
- Updatability
  - Argument is partially developed (because system is evolving)
  - But well-formed with open tasks and continuous update
    - linked anticipation and preparedness

Lifecycle Overview

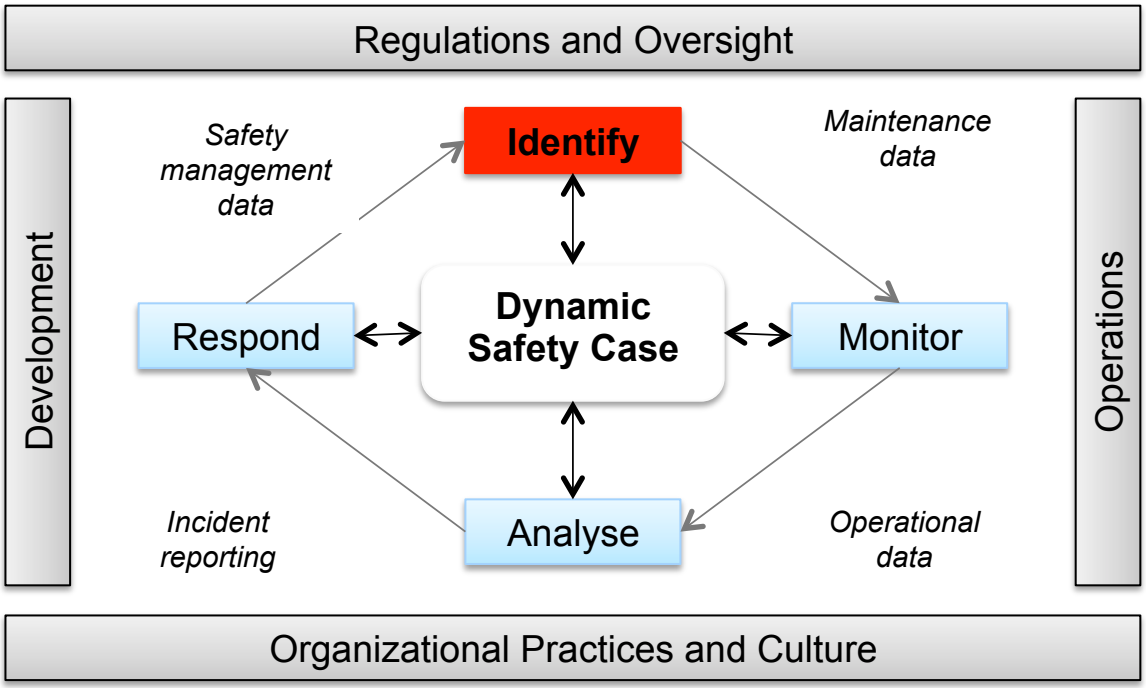


- Consideration of diverse factors
  - Development and Operations
  - Organizational practices and safety culture
  - Regulations

Plug the safety case into system operations

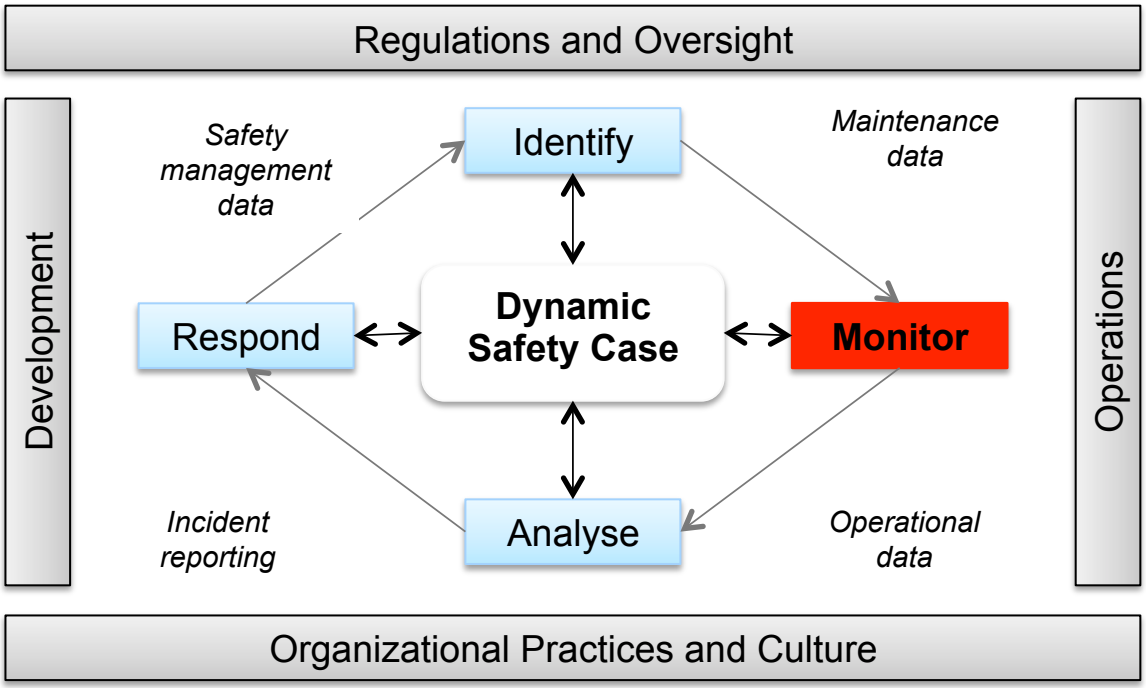
Identify

How can we decide on the most important subset of ADs?



- Sources of uncertainty in the safety case
  - i.e. assurance deficits (ADs)
  - Mapping ADs to assurance variables (AVars)
    - e.g., Environment and system variables
  - System/environment change → Argument change → AD Change

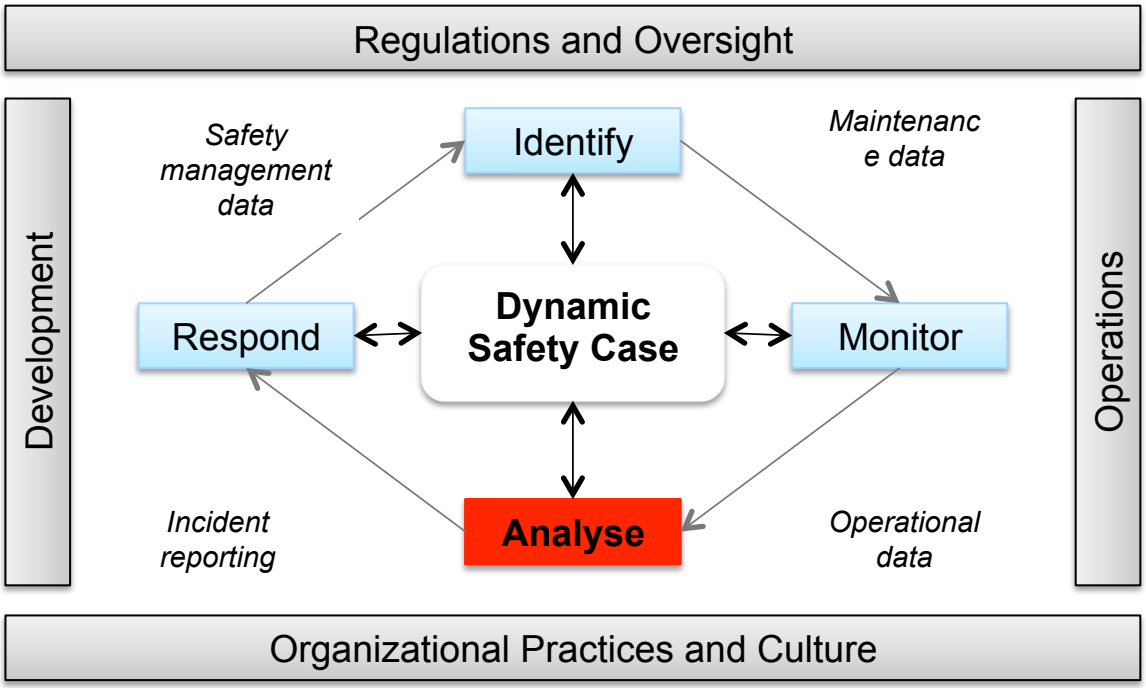
Monitor



- Data collection
  - Correspond to the underlying sources of uncertainty (AVars)
- Operationalize assurance deficits
  - i.e. Specify in a measurable or assessable way
- Relate to safety indicators
  - Leading / Lagging indicators

Analyse

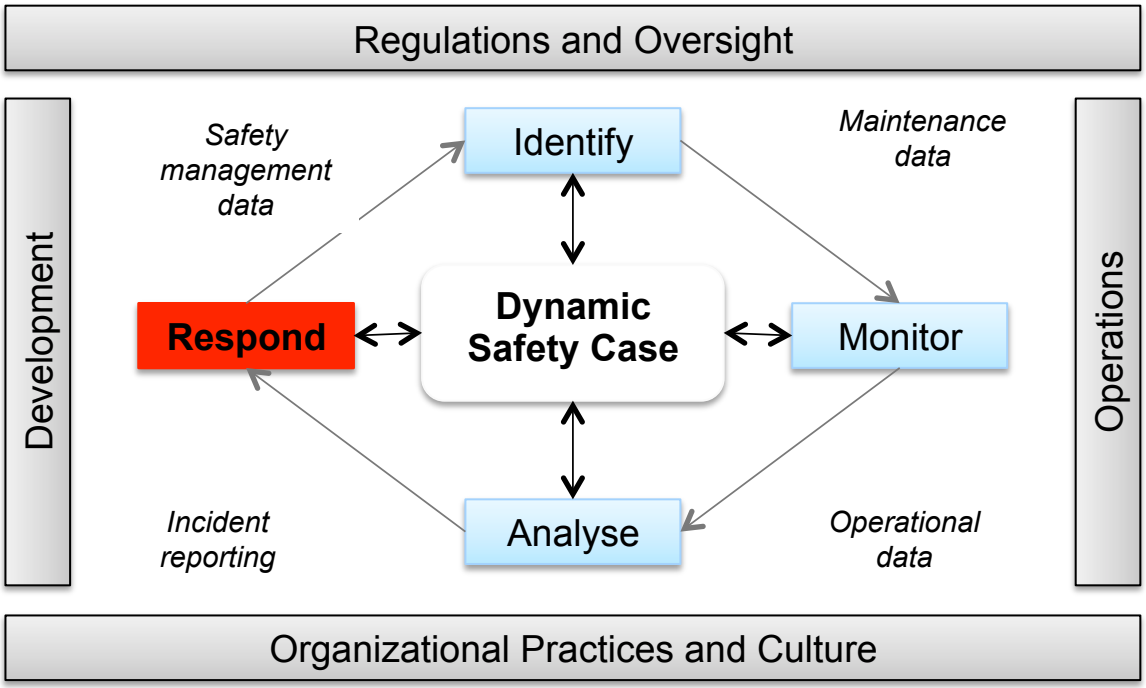
What can we learn from the world of AI and machine learning?



- Data analysis
  - Examine whether the AD thresholds are met
  - Define and update confidence in associated claims
- Interconnected Claims → Necessity to aggregate confidence
  - E.g., Bayesian reasoning?

Respond

Do we need a new theory for argument refactoring? Rule mining?



- Evolution
  - System / Environment change + DSC change, when necessary
- Basis of update rules
  - Impact on confidence of new data
  - Response options already planned
  - Level of automation provided
  - Urgency of response and communication

Dynamic Safety Case Elements

- Want to operationalise *through-life safety assurance*
  - Explicit argument structure + metadata
  - Confidence structure
  - Assurance variables
  - Monitors $(\mathsf{AVar}^* \rightarrow \mathsf{EnumVal} \mid \mathsf{ContinuousVal}) \times \mathsf{Period}$
  - Update rules $\mathsf{Condition} \rightarrow \mathsf{Action}^*$
  - Example:
    $$C[x] \Rightarrow \texttt{forEach}(y :: Q \mid A[x, y])$$
    - Remove a branch of the argument depending on an invalidated assumption

      $$\mathsf{not}(\mathsf{trafficDensity} < n) \Rightarrow \mathsf{forEach}(y :: \mathsf{solves}^* \mathsf{Contextualizes} \mid \texttt{replaceWith}(y, \texttt{empty}))$$

    - Create a task for an engineer to reconsider evidence when confidence in a particular branch drops below a threshold

      $$\texttt{confidence}(\mathsf{NodeX}) < n \Rightarrow \texttt{forEach}(E :: \mathsf{dependsOn}(E); \mathsf{traceTo}(\mathsf{NodeX}) \mid \texttt{createTask}(\mathsf{engineer}, \mathsf{inspect}(E), \mathsf{urgent}))$$

Related Work

- Formal foundation for safety cases
  - Work on automation and argument querying (Denney and Pai 2014)
- Measurement of confidence in safety cases
  - Confidence arguments modelled in BBN (Denney, Pai and Habli 2011/2012)
- Model-based assurance cases
  - Bringing the benefits of model-driven engineering, such as automation, transformation and validation (Hawkins, Habli and Kelly 2015)

Related Literature

In safety:

- Safety Management Systems
- Resilience engineering
- High Reliability Organisations
- Monitoring using safety cases
- …

In software engineering

- Models@runtime
- Runtime certification
- Conditional certification
- ….

One further consideration

What about unknown unknowns, i.e. total surprises?

Almost all theories in safety indicate that accidents are rarely total surprises

The information is out there but:

1. hard to find
2. complicated to analyse
3. given low priority
4. …

Thank you

Calinescu, Radu, Danny Weyns, Simos Gerasimou, Muhammad Usman Iftikhar, Ibrahim Habli, and Tim Kelly. "**Engineering trustworthy self-adaptive software with dynamic assurance cases**." *IEEE Transactions on Software Engineering* 44, no. 11 (2018): 1039-1069.