More than meets the AI: Can systems thinking leading indicators assist proactive safety in artificially intelligent systems?

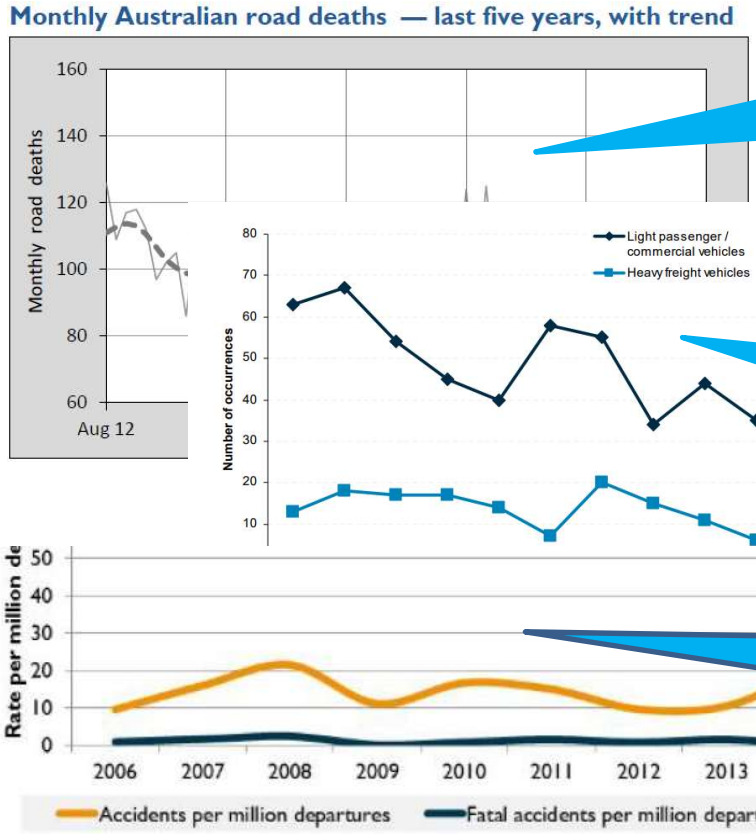
Eryn Grant
Human Factors Consultant
Acmena

The paper that aligns with this presentation is co-authored by Prof. Paul Salmon and Dr. Nicholas Stevens From the Centre for Human Factors and Sociotechnical Systems at the University of the Sunshine Coast.

Overview of Presentation

- The problem: incident rate in safety critical domains and the implications for AI
- Are there possible solutions in systems theory/systems thinking?
- What do we currently know about safety and accidents in sociotechnical systems and can this assist with proactive safety management?
- Case study showing how this concept may be applied for proactive safety
- Question: Can they benefit AI in assessing/maintaining systems safety

The Problem:
Incidents are not decreasing as they once were



A plateau effect on Australian roads (BITRE 2017)

Collisions between vehicles and trains has decreased but number of fatalities have not = increased severity of incidents (ITSR 2011)

An increase in aviation incidents from 9.5 per million dep. to 20 per million dep. in 2014 (ATSB 2016)

Acmena

Some hopes for AI

- Intelligent Assistance
- Decrease Workloads
- Increased Accuracy (i.e. diagnosis, weather forecasting)
- Freedom from human performance variability

The Problem:
Will AI introduce more or less risk into already ‘risky systems’?

Acmena

Can systems thinking help AI integration to achieve better safety outcomes?

- Safety and accidents are emergent properties of complex sociotechnical systems.
- Sociotechnical systems = how human, technical and social elements interact together to achieve a shared goal.

What do we currently know about safety and accidents in sociotechnical systems?

- The aim was to identify a core set of philosophies from systems thinking based accident causation models
- Review of over ninety published materials (books, journal papers, white papers etc.) using the authors below:

Jens Rasmussen | Erik Hollnagel | Charles Perrow | James Reason | Nancy Leveson | Sidney Dekker

What are the core philosophies of accident causation ?

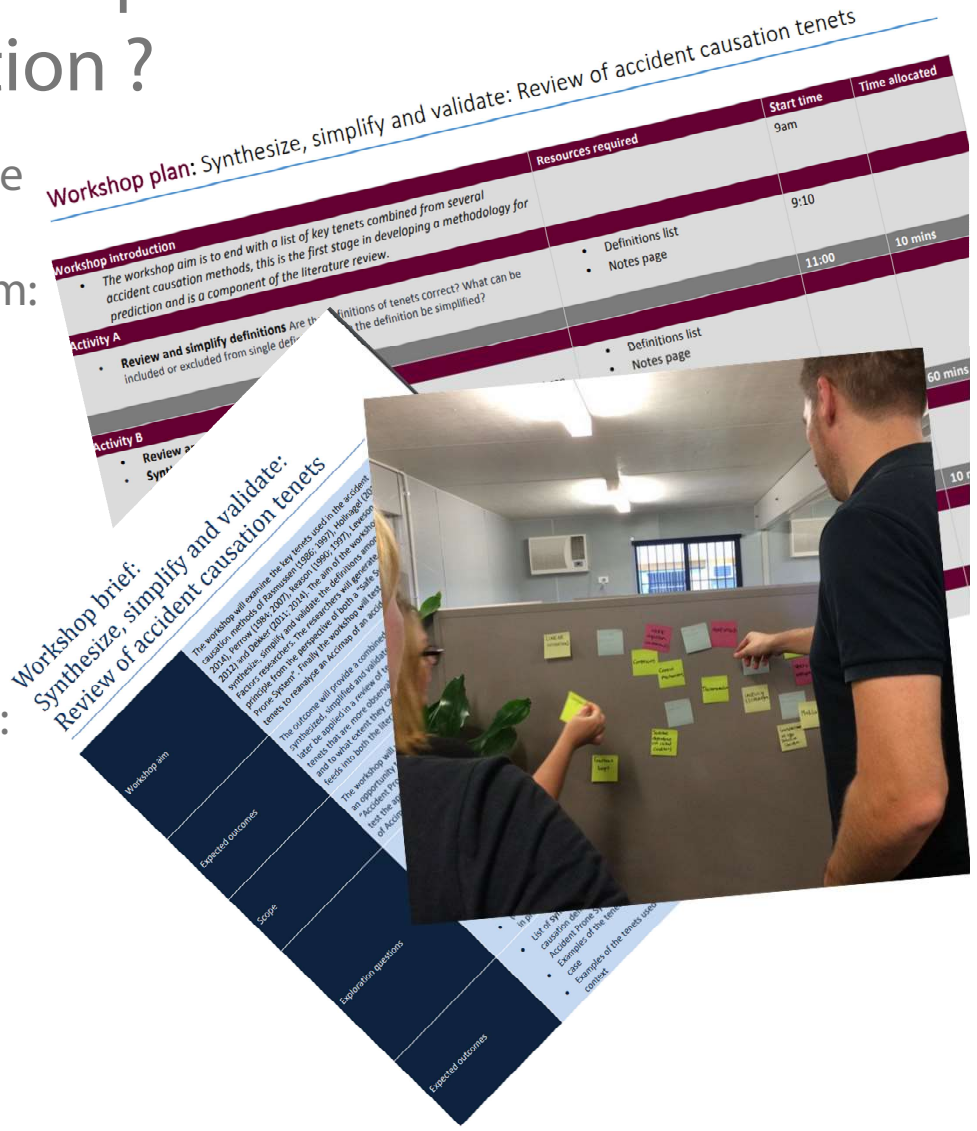
24 tenets were identified from the literature review process

A workshop was held to synthesize the them:

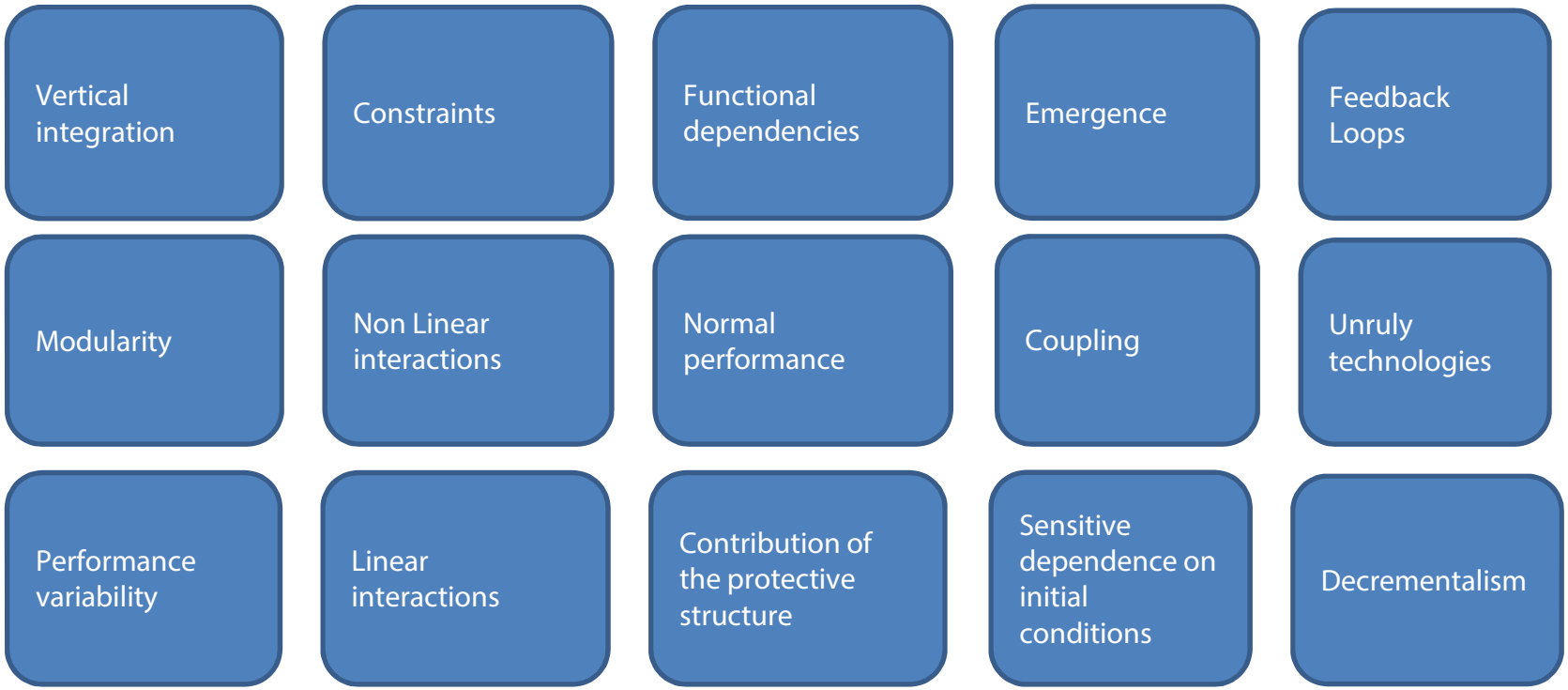
- Overlaps
- Exclusions i.e. redundancies, relevance to systems theory, context etc.
- Evidence of them in accidents and near misses

At the conclusion 15 tenets were recognized as the core philosophies of accident causation. Each were provided:

- ✓ Simplified definition
- ✓ Safe description
- ✓ Unsafe description



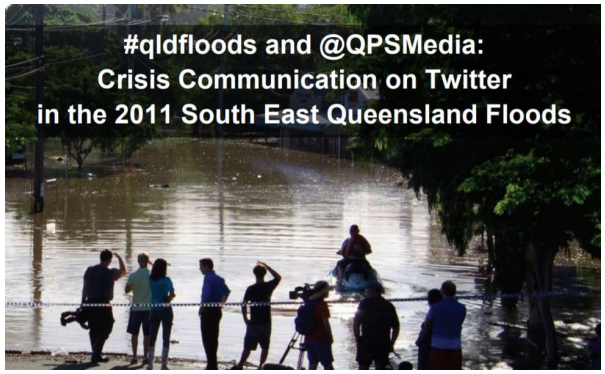
The core philosophies of accident causation, the systems thinking tenets



Acmena

Example: Unruly technology

Safe: Technology that supports adaptation through a mechanism that is beyond the scope of what is was designed for affording flexibility



Social media platforms have been found crucial in past crisis situations (e.g. Floods, Cyclones etc) by communicating emergency service info directly to those who need it most (Bruns et al. 2012)

Unforeseen behaviours or consequences of technologies.

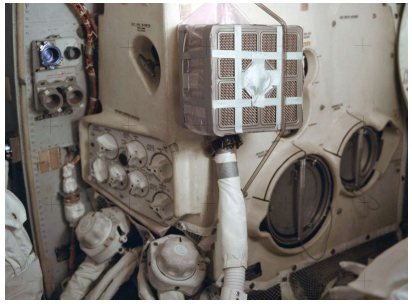
Unsafe: Technology that introduces and sustains uncertainties about how and when things may fail:

Pitot tubes measure airspeed. If for some reason they can't auto pilot will turn off configuring a plane to fly at alternate law

Acmena

Example: Vertical Integration

Safe: Decisions and actions at the higher levels filter down to lower levels and impact behavior. Information regarding the status of the system filters back up the hierarchy and influences higher level decisions and actions



Apollo 13: effective feedback across different levels of the system by prioritising information and needs between teams (Trotter et al, 2014)

Interaction between levels in the system hierarchy

Unsafe: Decisions and actions do not filter through the system and impact behavior on the front line. Information on the current status of the system is not used when making process decisions.

Walkerton E.coli outbreak. Information about the water quality did not filter through the system (Vincente and Christoffersen, 2006)

Acmena

Case Study: The Hawk Missile Simulation

Hawk pilot can not measure exact altitude

Flies 50ft above sea towards the frigate imitating a missile

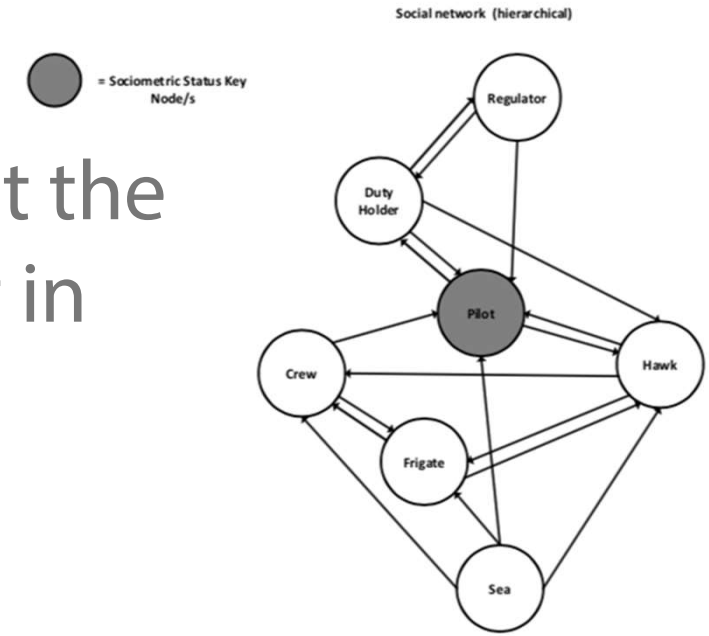
Acmena

Method: Assessing the system

- Model the system appropriately (Stanton & Harvey 2017)
- Define a set of rules/questions that allow the tenets to assess the system using the model.
- Assess if the system is operating within acceptable safety boundaries.

Results : Model the system

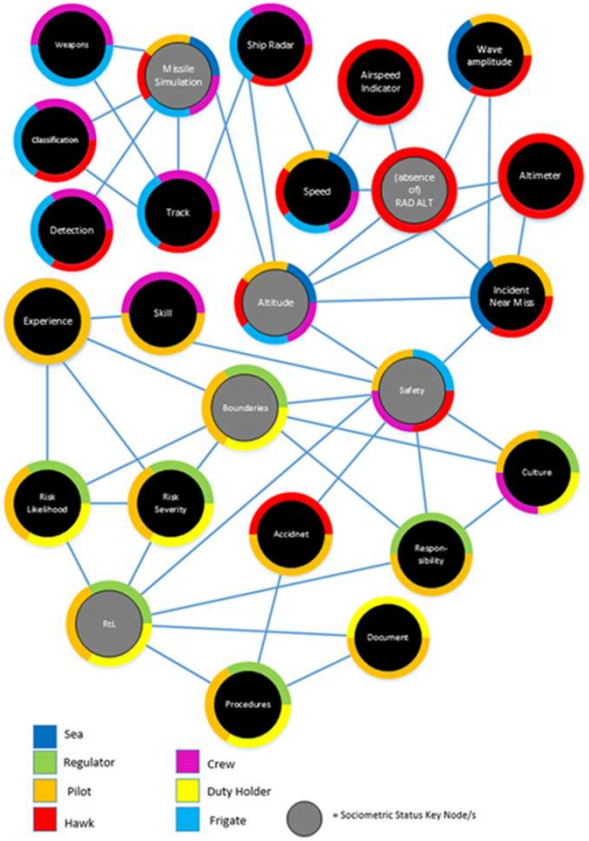
- Social Network
- Here you can see that the pilot is the key agent in the network



Stanton, N. A., & Harvey, C. (2017). Beyond human error taxonomies in assessment of risk in sociotechnical systems: a new paradigm with the EAST 'broken-links' approach. *Ergonomics*, *60*(2), 221-233.

Information Network

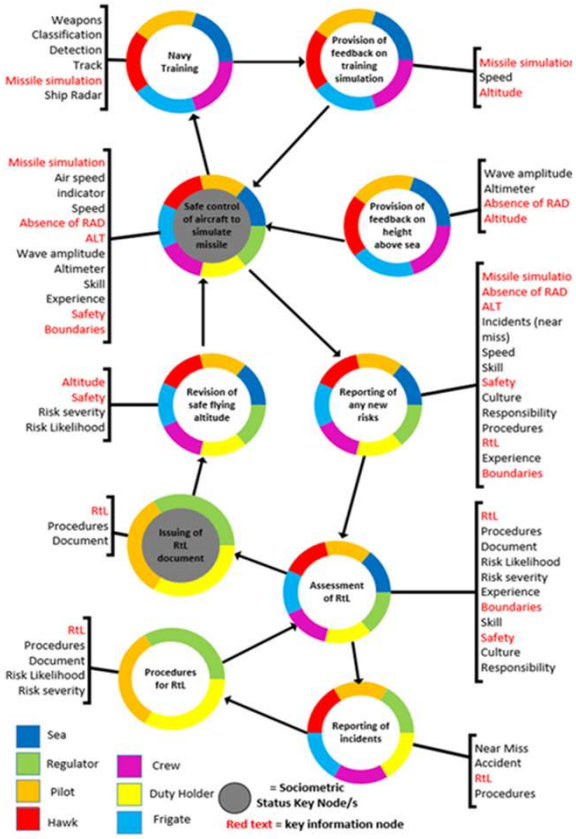
- The grey nodes define the key information required by the network.



Stanton, N. A., & Harvey, C. (2017). Beyond human error taxonomies in assessment of risk in sociotechnical systems: a new paradigm with the EAST 'broken-links' approach. *Ergonomics*, *60*(2), 221-233.

Composite network

- Based on the task network. This differs Stanton and Harvey (2017)
- Nodes show tasks with associated agents (coloured lines) and information
- Grey nodes indicate the key tasks in the system.



Stanton, N. A., & Harvey, C. (2017). Beyond human error taxonomies in assessment of risk in sociotechnical systems: a new paradigm with the EAST 'broken-links' approach. *Ergonomics*, *60*(2), 221-233.

Acmena

Applying the rules to the EAST model of the Hawk simulation system

- For example using the tenet Constraints

Constraints	Definition: Influences that limit the behaviours available to components within a system.	Safe: Specific constraints introduced to control hazardous processes	Unsafe: Restricts appropriate performance variability

Applying the rules to the EAST model of the Hawk simulation system

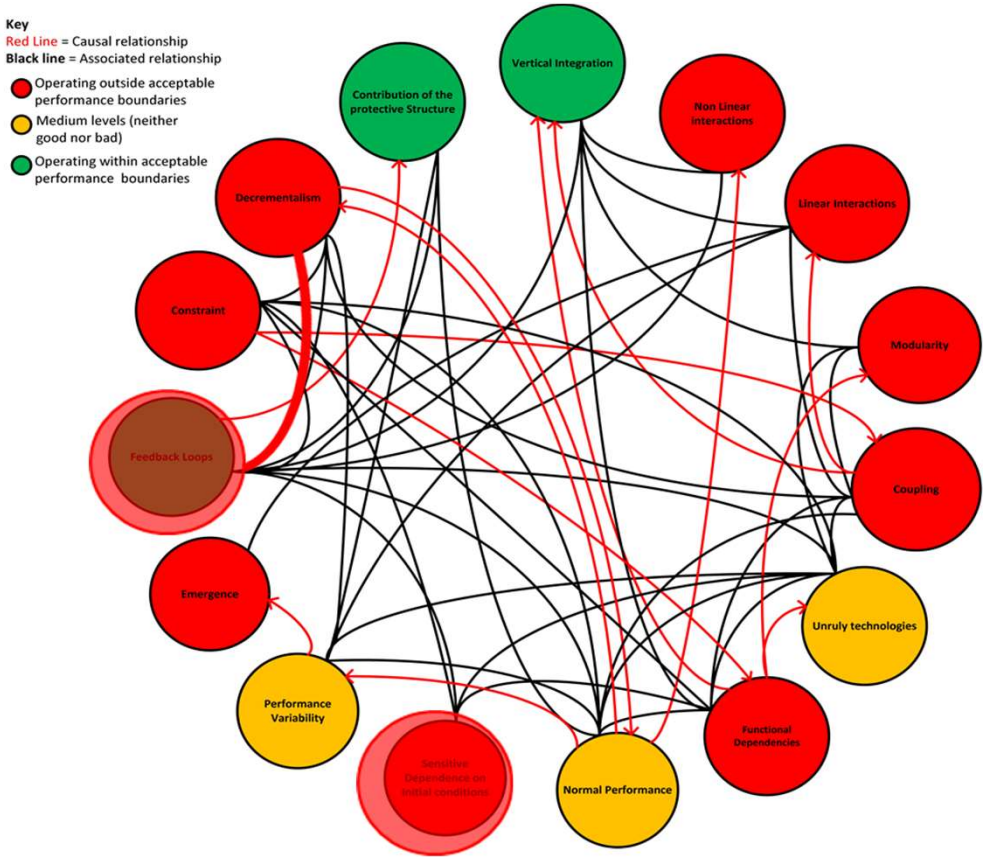
- For example using the tenet Constraints an assessment is made by asking :
  - Are time constraints placed on tasks?
  - Are there physical objects that restrict t task can be performed?
  - Are fixed actions, operations or functions present?
  - Then rate the tenet.

Yes

The Hawk jet, absence of RAD ALT, the sea

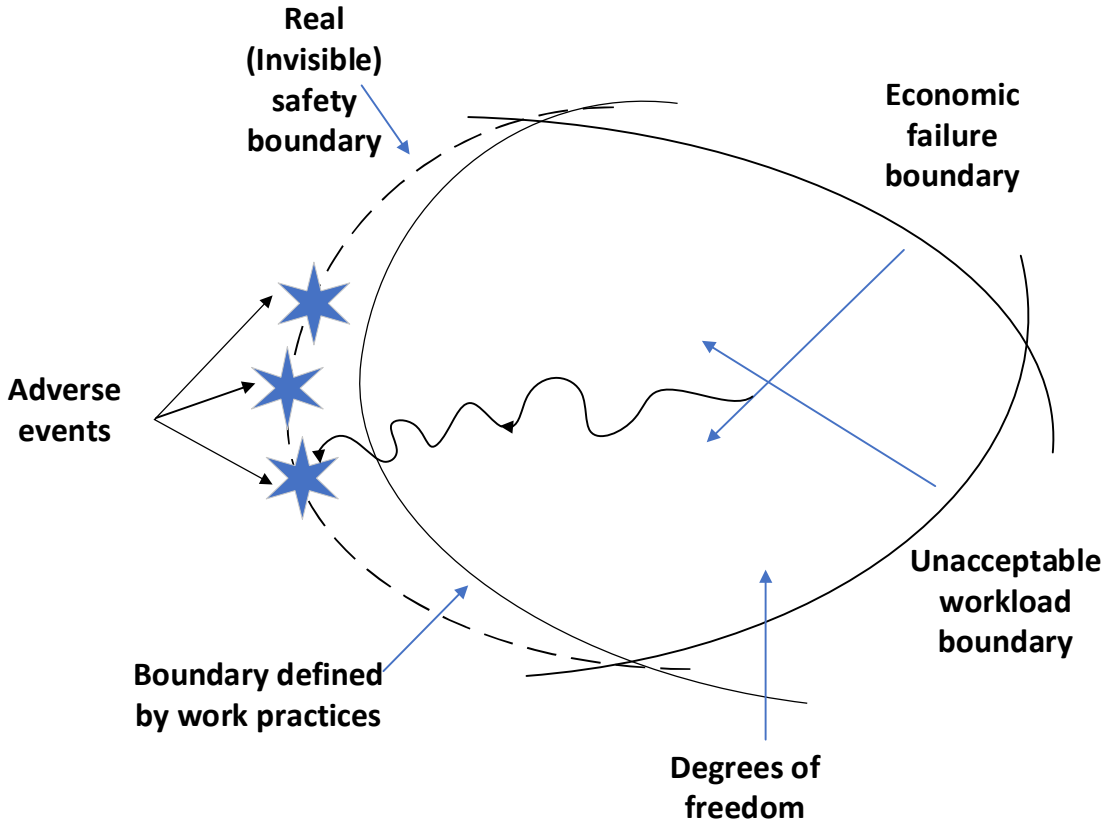
The missile simulation (height above sea, speed). How many simulation flights required for training?

Assessing the system



Acmena

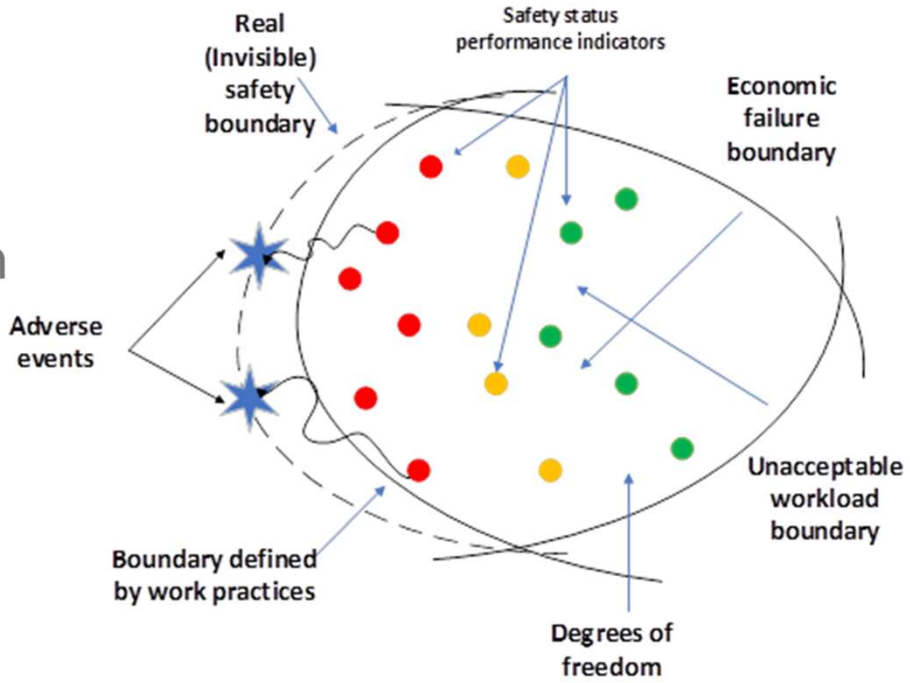
What could this mean for AI?



Based on Rasmussen's Risk Management Framework (1997) The boundaries of acceptable performance

What could this mean for AI?

- AI systems may use the information from the systems thinking tenets to monitor and learn from system performance, indicating when a system may be to closer to performance boundaries and unwanted outcomes



What could this mean for AI

- For example: In the hawk system missile simulation heights are set by a regulator – mostly as a reaction
- …..But do they need to be? Develop a way to constantly monitor the system using the tenets and set boundaries based on current needs.

Conclusions

- It is uncertain whether introducing AI into already 'Risky Systems' will produce safer outcomes.
- The core philosophies of systems thinking may help as leading indicators that provide a way for AI to learn prior to incidents occurring

Questions

Contact details:

eryngrant@acmena.com.au

@What_the_HF

Eryn Grant