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

Eryn Grant[1,2], Paul M Salmon[2] & Nicholas J Stevens[2]

[1] Acmena PTY LTD 24/86 Brookes Street, Fortitude Valley, QLD 4006
[2] Centre for Human Factors and Sociotechnical Systems
University of the Sunshine Coast, 90 Sippy Downs Dr, Sippy Downs QLD 4556

eryn.grant@acmena.com.au
psalmon@usc.edu.au
nstevens@usc.edu.au

Abstract

Whilst accident analysis is an accepted approach to safety management, it is reactive in nature and often requires the occurrence of adverse events to learn appropriately. For the introduction of Artificial Intelligence (AI) systems, this is problematic, as learning may well rely on things going wrong to build on future system safety. AI represents some significant complexity and the risk of unintended consequences and system behaviours makes it difficult to model and anticipate how safety issues might arise. This paper suggests that by integrating what is currently known about accident causation with AI technologies, it may assist AI to achieve optimal learning without adversity.

A proactive approach to safety using AI may be achieved by monitoring 'normal performance'. However, there are few methods supported by safety science that provide specific support on the identification of the conditions that could create accidents and safety compromising events that integrate whole systems properties without focusing on end errors. The research presented in this paper is a response to this capability gap. It presents the core tenets of accident causation that may be applied as performance indicators to proactively assess complex work systems normal performance and provide learning opportunities for AI. Implications for future research are also discussed.

Keywords: Systems Thinking, Artificial Intelligence, Human Factors and Ergonomics, Safety Boundaries, Leading Indicators.

1 Introduction

Increasing system safety through reducing adverse events remains the major challenge for safety science (Hollnagel 2018, Salmon et al. 2017, Waterson et al. 2017, Dekker & Pitzer 2016, Carayon et al. 2015). From the perspective of systems thinking, safety and accidents are the result of emergent behaviours in systems where interrelated components work to achieve systems goals (Stanton et al., 2012, Leveson 2004). The complexity of systems and the environment in which they operate means the process of safety is never straightforward or linear, but instead is underpinned by a complex web of relationships and behaviours between humans, technology and their environment (Underwood & Waterson, 2014).

Retrospective analysis has enabled the identification of incident characteristics to learn from past events and support accident prevention (Leveson & Dekker, 2014, Hancock & Chignell, 2018). However, data collected on incidents overtime has shown a plateau (or indeed an increase) in multiple safety critical domains, including road, rail and aviation (Grant et al., 2018, Leveson, 2012; Salmon et al., 2016). This suggests that accident analysis is underperforming, and traditional approaches may have reached a saturation point or are inadequate to solve contemporary complexities of such systems. That is, they can no longer inform appropriate measures to reduce recurring incidents of the same nature (Maurino et al. 2017).

When introducing novel technologies into safety critical domains that experience variability, safety is a key concern. Artificial intelligence (AI) is one such technology. Whether general AI, deep learning or machine learning, at some level AI is present in daily life. Developments in AI are trickling down into the 'everyday' (e.g. Apple's Siri, Google's Alexa) and in many cases users are not fully aware of how AI is a participant in their daily routine (Porcheron et al., 2018). AI was broadcast mainstream when computer intelligence conquered a human in a game of chess in 1997 (Fogel, 2006). Since this time, AI has advanced in capability and is able to learn thousands of years of human knowledge in less than a day (Singh et al., 2017). Yet it may not always require human intervention to learn, AlphaGo Zero, with reinforcement learning which means it is free of human guidance. Within four hours, AlphaGo Zero mastered the ancient Chinese board game, 'GO' with no human help beyond being told the rules. It succeeded to win against the best human players and superseded the former AI master, AlphaGo (Singh et al., 2017).

If the excitement of AI is to be believed, daily life will change significantly (Pelton, 2019). Full autonomy is promised for automobiles offering freedoms from physical and mental workloads and with it the promise of more predicable safety outcomes (Herrmann, 2018). Better health results for those with disease, with the potential for inexpensive medical diagnosis that is more accurate, but also more available and less expensive with the capacity to reach remote and underrepresented populations (Mayo & Leung, 2018).

However, a question remains where the seduction of such promises may be blinding us to other consequences of AI and how it may affect our systems in unimagined ways (Hancock, 2018). First, in what ways are the promises of AI balanced with the current realities of systems that it will be introduced. For example, incidents where cars run over pedestrians as an algorithmic choice have made already risky systems unruly (Rice 2019). If medical advances mean populations live longer, what systems are in places to manage population growth and

aged care, employment and basic resources such food, water and housing?

Second, if systems are variable and unpredictable will AI introduce less or more of the same? And what resources are there to monitor this and any other variability caused by its introduction? Safety in this context is not simple, as Woods & Hollnagel (2017:4) state, "in a world of finite resources, of irreducible uncertainty, and of multiple conflicting goals, safety is created through proactive resilient processes rather than through reactive barriers and defenses.". That is, it would be irresponsible, with the many variabilities that are unknown, for safety management to be reactive to the unwanted consequences of AI. Safety management in this context must be based on a proactive strategy, rather than a reactive one (Hollnagel, 2017).

In raising these questions, this paper aims to describe how AI may benefit from integrating the core philosophies of accident causation, derived from leading systems theory-based accident models. This may assist in managing safety proactively, whereby AI "learning" is performed using a set of "leading indicators" derived from core safety philosophies. In doing so the paper presents an overview of the core philosophies of accident causation to show how performance boundaries may be identified as systems thinking-based leading indicators to assess the safety boundaries of complex work systems.

2 Systems thinking and accident causation

Utilising the theoretical perspective of systems theory, safety and accidents are emergent properties of complex sociotechnical systems. Sociotechnical systems refer to the interaction between people and technology within a broader organisational framework of a system (Trist 1981). While always tacitly present, sociotechnical systems are how human, technical and social elements interact together to achieve a shared goal (Walker et al., 2008). As a system, interrelated objects, actors and functions work together collectively (Leveson, 2013). These interactions shape behaviour in systems that leads to either safe or unsafe outcomes (Underwood & Watson, 2013).

Accident causation models underpinned by systems thinking principles describe how interactions create emergent outcomes. According to contemporary models, accidents occur because of multiple interacting parts that cannot be attributed to any one person or thing. Accident causation models underpinned by systems thinking do not seek an end error event or attribute blame toward a person or component as the primary cause of the accident (Dekker 2011). Instead, accident models underpinned by systems thinking describe accidents through the interactions and causally related factors across a system that contributed to an unwanted outcome.

2.1 Systems thinking based accident causation models

Many accident causation models underpinned by systems thinking exist. Despite sharing the same systems thinking principles or philosophies, the models characterise systems differently. Rasmussen's Risk Management Framework (1997) is a functional abstraction model that describes accidents by demonstrating the existence of instability in the system, caused by a lack of vertical integration, which ultimately means a loss of control. Levesons (2004) Systems Theoretic Accident Model and Processes (STAMP) primarily views accidents as a problem of controls and constraints. Controls and constraints are imposed on a system to limit system behaviour and ensure it operates within safe boundaries. Accidents occur when these controls and constraints fail to control components and their interactions with one another. Other models such as Functional Resonance Analysis Method (FRAM; Hollnagel 2012) do not explain systems hierarchically or by abstraction. A FRAM analysis outlines the mutually coupled or dependent functions of a system to explain variability. Some offer no definitive model and rather explains sets of philosophies (e.g. Dekker 2011) or explain between system components in specific ways (e.g. Perrow 2007). Despite their modelling differences they all are based on the same premise; unwanted outcomes are emergent properties that arise because of interactions between components in a system.

3 Applying the systems thinking tenets: leading indicators to assess performance boundaries

The aim of this paper is to describe how AI may benefit from integrating the core philosophies of accident causation derived from leading systems thinking-based accident causation models. In doing so the paper presents an overview of an assessment of performance boundaries in a complex work system presented in Stanton & Harvey (2017). From this assessment the paper proposes how AI may benefit from systems thinking-based leading indicators that may afford proactive learning of systems specific vulnerabilities.

Performance boundaries in a system are constraints that the work system must maintain to operate successfully such as economic or workload boundaries (Rasmussen 1997). Many degrees of freedom exist within performance boundaries, where normal changes to work conditions can occur. Overtime variations to work may place pressure on performance boundaries to maintain constraints (Rasmussen 1997). When this occurs, a system may migrate to functionally acceptable performance boundaries and, if crossed the results may be irreversible (i.e. accident occurrence).

The current paper describes an overview of how proactive safety may be assessed using a set of systems thinking tenets, that represent the core philosophies of the five leading systems theory-based accident causation models. The systems thinking tenets define performance using fifteen core properties of systems thinking that are related to accidents and safety in sociotechnical systems. Finally, the paper will propose the tenets communicate systems specific vulnerabilities and afford an opportunity whereby AI may holistically apply them as a learning strategy for proactive safety management.

3.1 The systems thinking tenets

Previous review and synthesis of contemporary accident causation models has enabled the identification a set of common tenets that represent the core philosophies of the five leading models. The models were selected on

the basis that they were underpinned by systems thinking and had a measure of importance when applied within academic safety literature identified by citation records (Scopus 2016). The models assessed included the Systems Theoretic Accident Model and Processes (Leveson 2004), Risk Management Framework and AcciMap technique (Rasmussen 1997), The Drift into Failure model (Dekker 2011), Functional Resonance Analysis Method (Hollnagel 2012) and Normal Accident Theory (Perrow 2011). From these, fifteen systems thinking tenets were identified that represent the core philosophies of accident causation. Grant et al. (2018a) describes the process used to identify and validate the tenets. Further studies have shown how they are present in accidents and systems ergonomics methods that best identify them as behaviors of sociotechnical systems (Grant et al. 2018b; Grant et al. 2017). An important contribution of the tenets is their capacity to define safe properties that contribute to "normal" operation in addition to unsafe properties that contribute to accidents. For this reason, both safe and unsafe characteristics of the systems thinking tenets are presented (see Table 1). Grant et al., (2018) argue that the tenets provide a useful framework for identifying pre-accident conditions and likely emergent accidents in complex work systems.

3.2 The context: Hawk Jet "missile simulation" naval training exercise

The systems thinking tenets were applied to assess the safety status of a complex work system using the analysis available in Stanton & Harvey's (2017) study on a naval training exercise.

Stanton & Harvey (2017) outline a novel risk assessment approach using a systems ergonomics method; Event Analysis of Systemic Teamwork (EAST; Stanton et al. 2013; Walker et al. 2010). The risk assessment involved breaking information links to identify the significance of information loss within the system under investigation. However, for the purposes of this paper, the source of information was also useful to assess performance boundaries of the system by using the systems thinking tenets to interpret the EAST networks as a model of the system.

EAST follows a 'networks of networks' approach in which three interlinked network-based models are used to describe and analyse activity. EAST describes system functions through three networks (task, social and information) showing how networks interrelate in the organisation of the system.

Task networks are used to describe the goals and subsequent tasks being performed within a system. Social networks are used to analyse the organisation of the system and the communications taking place. Information networks show how information and knowledge is distributed across different agents within the system (Salmon et al. 2014). Task, social and information networks are finally combined to create a composite network revealing the complexity of the system and the relationship between the three networks (Stanton 2008, Stanton et al. 2017).

3.2.1 Description of the Hawk missile simulation

The Hawk missile simulation involves a jet simulating a sea skimming missile. The Hawk must maintain a low altitude yet is missing the technology that provides precise altitude measurements (RADALT). Yet the principle simulating device (HAWK) does not have that technology but must mimic the flight profile of the missile.

The Hawk missile simulation involves a highly experienced pilot flying a Hawk jet at high speeds and at low altitudes (50ft) above water (Stanton, Harvey & Allison, 2019). The Hawk is flown towards a frigate where the crew on board will "detect" the Hawk as a missile and simulate mitigation strategies of control (for example a 'target and fire' sequence). For obvious reasons this training cannot be performed with a real missile, however as the threat exists for Navy crews in various combat situations, appropriate crew training is necessary. The Hawk jet simulates real-life speeds and height above sea of a missile attack, affording frigate crews the best possible substitution for a training sequence on the sea. A significant problem for the pilot of the Hawk jet is maintaining accurate speed and height above sea to adequately simulate a missile. This is due to an absence of instruments that provide precise altitude readings (a Radio Altimeter or RAD ALT). The pilot must use their own skill and judgement in conjunction with a traditional barometric altimeter to maintain an altitude of 50ft above sea level. The altimeter reading is not precise but provides enough guidance for experienced pilots to estimate their altitude. This, combined with the high speeds and volatility of the sea over which they must fly, means retaining a Risk to Life (RtL) assessment that is practicable is extremely complicated.

3.3 Method

To perform the safety status assessment the information from EAST models constructed by Stanton & Harvey (2017) were used. The systems thinking tenets were applied to the EAST networks to assess the model of the system and determine how safe the system was compared to an "ideal" level of tenet performance.

3.3.1 Modelling the Hawk missile simulation system

EAST Networks

The information relating to the EAST task, social and information networks were used from the analysis available in Stanton & Harvey (2017). Networks were re-drawn using this information to meet the requirements of the tenet analysis that is described in this paper. Stanton & Harvey (2017) provide details on how the EAST networks were assembled including data collection and subject matter expert (SME) validation of the EAST network models.

Network Analysis

The task, social and information networks were analysed using the following metrics:

Network Density

Network density is a whole network metric calculated on the interconnectivity of the network. This calculation is a value between 0 and 1, where 0 represents a network with no connections between nodes and 1 represents a network where every node is connected to all other nodes. Values closer to 0 represent networks that are loosely coupled and those closer to 1 represent networks that tend to be more tightly coupled (Kakimoto et al. 2006).

Sociometric status

Sociometric status is an individual node metric that is a comparison between how active a node is in a network and how many nodes exist in the network (Haughton et al. 2006). A key node is selected on the basis that it has a higher sociometric value than the mean + standard deviation, as this shows the node is more related to others in the network.

3.3.2 The system thinking tenet assessment process

A set of questions or rules were designed to assist the lead author apply each tenet to the EAST model and then assess the performance of each tenet with a safe or unsafe "value". This required classifying performance relative to a tenet in terms of what may be considered "optimal" for the system and then querying if this performance was reached using the EAST model as an example of system operation. For example, using the tenet "constraints", specific constraints may exist in the system that are identified via the EAST model (e.g. physical objects, temporal or functional). Using the composite network in Figure 2, constraints in the system were identified by assessing the network using the following questions:

1. Are time constraints placed on tasks?
2. Are there physical objects that restrict the way a task can be performed?
3. Are fixed actions, operations or functions present?

If constraints do exist, a determination could be made on the way they provide a safety value or safety restriction to the overall work system. Constraints within the Hawk missile simulation system could then be compared between what was considered ideal (which is determined by an assessment using the tenets based on subject matter expertise of the system) and how constraints are realised through system operation modelled by EAST.

3.3.3 Assessing performance boundaries

Materials

To perform the tenet analysis three EAST networks (task network, social network and information network) in conjunction with a pre-designed set of rules were used to provide the lead author a means to assess the activity of each tenet when applied to assess the system modelled by EAST.

Procedure

A set of questions were designed to assist in evaluation of the model of the system in terms of optimal performance as described in section 3.3.2.

Using the answers to these questions, information was recorded through the rating structure of green, amber or red. In summary the rating system followed these general principles:

- If assigned a green rating, the tenet performed optimally in the system when responding to the rules.
- If assigned with an amber rating, the tenet performed to a medium level of acceptability in relation to the tenet rules.
- If assigned with the red rating, the tenet did not perform acceptably in relation to the tenet rules.

An additional inclusion was the lead authors interpretation of any causal or related tenets when applying each rule (see section 3.3.4 for more details on this process). Once completed the results were reviewed by the second and third authors with experience in applying Human Factors and systems ergonomics methods (Salmon et al. 2017, Stevens 2018).

3.3.4 Identifying causal and associated tenet relationships

Interconnected behaviours are an important component of systems thinking. Relationships were identified based on two properties;

- If one tenet influenced the status of another it was identified as a causal tenet.

As an example of a causal tenet relationship the analysis suggests that Normal Performance is causally tied to Decrementalism, whereby work practices may be done differently to prescribed rules. When combined with constant small changes to tasks overtime (that are acceptable or sanctioned) this can lead to large gaps between work as imagined at the higher levels of the system and how it is done at the lower levels. If adverse events occur, they draw attention to the large gap that has formed over time between how tasks were imagined at the higher levels of the system and how they were achieved at the operational level.

- If tenet properties were present or maintained in the analysis of another tenet (but not causally related) they were identified as associated

As an example of an associated tenet relationship, the analysis suggests that Decrementalism could potentially be present in incident reporting, an associated tenet is Feedback Loops because the incident report is a mechanism to provide system feedback.

4 Results

4.1 East networks

Task Network

The composite EAST network representing the combined task, social and information networks is presented in Figure 1. Key tasks include 'safe control of aircraft for missile simulation' and 'Issuing of RtL documentation'. Figure 1 is also a composite network, indicating what social agents and information related to each task. The key information nodes are highlighted in red. The network has a density of 0.13 which indicates a relatively loosely coupled network.

Table 1. The System Thinking Tenets (Grant et al., 2018)

	Definition	Safe	Unsafe
Vertical Integration	Interaction between levels in the system hierarchy	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	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
Constraints	Influences that limit the behaviours available to components within a system.	Specific constraints introduced to control hazardous processes	Restricts appropriate performance variability
Functional dependencies	The necessary relationships between components in a system.	Relationships between functions are expected and sustained	Dependencies that are not wanted or expected
Emergence	An outcome or property that is a result of the interactions between components in the system that cannot be fully explained by examining the components alone.	Emergent behaviours that support the goals of the system	Behaviours that undermine the goals of the system
Normal performance	The way that activities are actually performed within a system, regardless of formal rules and procedures	Behaviours are flexible enough to cope with adverse conditions	Behaviours cannot cope with the unfolding situation
Coupling	An interaction between components that influences their behaviour; both tight and loose interactions	Tight: connections between components are evident Loose: recovery from disturbances in the system is possible	Tight: Cascading failures when one component breaks down Loose: Loss of control of regulating behaviours. Too much independence. Duplication of functions leading to inefficiencies.
Non-Linear interactions	Interactions are complex relationships between components where the outcome is not predictable	Allows for adaptations in the system.	Inconsequential events have large effects, cannot predict the effect of changes
Linear interactions	Direct cause effect relationships between components where the outcome is predictable.	Predicable and dependent	Interactions are predefined and fixed with no allowances for adaptations
Modularity	The organisation of a system where sub systems and components interact but are designed and operate largely independently of each other.	The system is resilient to breakdowns, replacement or substitutions of components and organisation of sub systems can be easily made	The system is tightly integrated and complex, substitutions cannot be made
Feedback Loops	Communication structure and information flow to evaluate control	Feedback is received on system breakdowns allowing control of hazards	Communication structures are not in place to provide or receive system feedback.

	Definition	Safe	Unsafe
	requirements of hazardous processes		
Decrementalism	Small changes in normal performance that gradually result in large changes.	Complex systems need to adapt, small adaptations are required to maintain optimisation	Constant small organisational changes create conflicts and pressure
Sensitive dependence on initial conditions	Characteristics of the original state of the system that are amplified throughout and alters the way the system operates (interconnected webs of relationships).	Mechanisms for monitoring changes are available	No understanding of initial conditions and their influence on the system
Unruly technologies	Unforeseen behaviours or consequences of technologies.	Technology that supports adaptation through a mechanism that is beyond the scope of what is was designed for affording flexibility.	Technology that introduces and sustains uncertainties about how and when things may fail
Performance variability	Systems and components change performance and behaviour to meet the conditions in the world and environment in which the system must operate.	Performance varies to meet the needs of changing conditions	Performance does not change when conditions change
Contribution of the protective structure	The organised structure and system control that are intended to optimise the system, instead they do the opposite.	Protective structures are effective, flexible and adaptable in maintaining controls	Protective structure inhibits performance variability. Introduces new tasks that do not contribute to the goal. Unnecessary controls

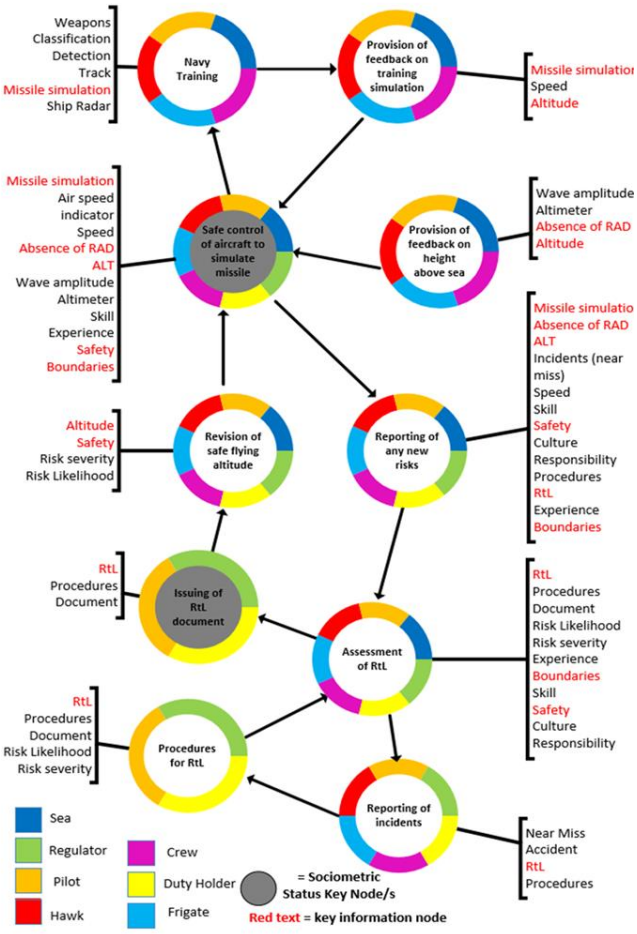

Figure 1. Composite EAST network representing the combined task, social and information networks for the Hawk missile simulation system.

Social Network

The social network is represented by the coloured markings around the task nodes in Figure 1. The pilot was identified as the key node within the social network with a higher sociometric value. This indicated the pilot was more connected in the network than any other "agent". The network has a density of 0.13, which indicates a very loosely coupled social network.

Information Network

The information network is represented by the information connected to tasks in Figure 1. Key nodes include 'Missile Simulation', '(absence of) RAD ALT', 'Altitude', 'Safety', 'Boundaries' and 'Risk to Life'. The network has a density of 0.45, which indicates a moderate level of coupling between information nodes.

4.2 Results of safety status performance

The results suggest that three tenets (n=3) were rated as green: Vertical Integration, Feedback Loops and Contribution of the Protective Structure. Four tenets (n=4) were given an amber rating, including Unruly Technologies, Sensitive Dependence, Normal Performance and Performance Variability. Red or negative results were found for the eight remaining tenets (n=8). These tenets included Coupling, Functional Dependencies, Modularity, Linear Interactions, Non-linear Interactions, Decrementalism, Constraints, and Emergence.

4.2.1 Causal and related tenets for the Hawk missile simulation case study

Figure 2 is a representation of the causal and related tenets found in the Hawk missile simulation. The red lines with arrows indicate the direction of causal relationships (i.e. one tenet behaviour impacts the outcomes of another). No causal relationships were repeated, however the most frequently identified causal tenets in the Hawk missile simulation system were Normal Performance and Functional Dependencies both identified three times (n=3). As shown in Figure 2 Normal Performance is causally related to Non-linear Interactions, Performance Variability and Decrementalism. Functional Dependencies is causally related to Vertical Integration, Modularity and Unruly Technologies. Other causal tenets include Constraints and Coupling both identified twice (n=2); Feedback Loops, Decrementalism and Performance Variability all identified as causal tenets one time (n=1). No significant relationships were noted as repeated relationships however the tenets Feedback Loops and Unruly Technologies shown to be associated eight times (n=8) to other tenets. This was followed closely by Normal Performance associated seven times (n=7).

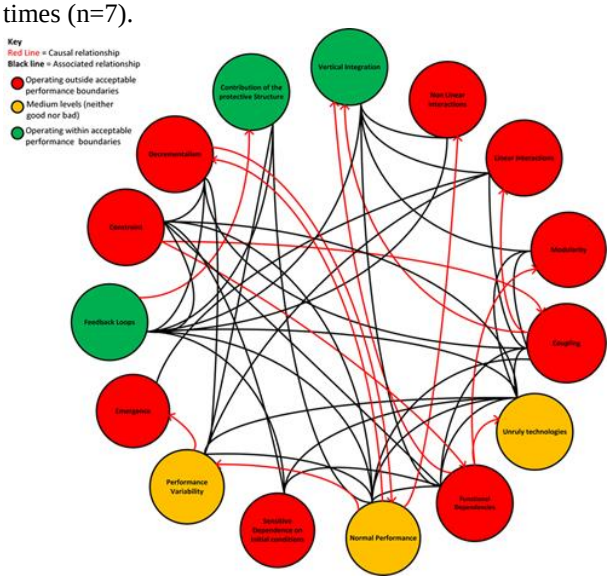

Figure 2. Safety status and relationships represented between tenets in the Hawk Missile simulation system.

5 Discussion

The analysis results suggest that the Hawk missile simulation has high levels of variability in the system that could potentially impact safety. The system relies on feedback from its own performance where this may not always occur. Feedback Loops were shown to be highly associated to other tenets in the system. While Feedback Loops were rated as green, suggesting it performs at optimal levels, the analysis results also suggest that when combined with other tenets, this may contribute to unwanted information or performance feeding through the system.

For example, the combination of Feedback Loops and Decrementalism may afford a positive Feedback Loop of

unwanted task changes (such erosion of the incident reporting task, where over time reporting of incidents and near misses does not occur). In such situations, Feedback Loops communicate that the system is incident free, when this may be incorrect, affecting how the RtL assessment is ultimately measured.

The analysis presented in Figure 2 suggests that Unruly Technologies were highly associated to other tenets in the system, most likely due to the use imprecise equipment used to measure altitude. For example, a technology that is present but is not suitable for the purpose of the flight activity. The analysis showed the inability to measure altitude placed pressure on most other activities in the system.

5.1 Performance boundaries of the Hawk missile simulation system

5.1.1 A loosely coupled task network

The density rating of the task network was assessed as 0.13 which is indicative of a very loosely coupled network. Loose Coupling in task networks may indicate task sequences that are more difficult to maintain. This was also found in Non-linear Interactions showing that no protections were in place to inhibit tasks from occurring out of sequence. For example, the regulated As Low as Reasonably Practicable (ALARP) rating for the RtL risk assessment is based on known incidents and hazards; if tasks responsible for information relating to these have not been completed or completed out of sequence the RtL risk assessment may not be current. This means regulations based on incidents (such as safe flying altitude) may not be based on the most current information. One protection from this is the higher Coupling of the social network, along with key information pathways showing medium levels of density for Vertical Integration (the density is 0.45). This would require that agents in the social network identify that information required for tasks is lacking or is not current, allowing time to complete tasks that have not been done.

5.1.2 The distribution of information in dynamic settings

As military settings are dynamic and interactive, overtime command and control structures have shifted from a traditional hierarchical top down approach to one where leadership is more distributed through teamwork (Ramthun and Matkin 2014). Given this more fluid distribution, determining positive information flows, feedback and control structures are operating soundly is essential. For the Hawk missile simulation, Vertical Integration, Contribution of the Protective Structure and Feedback Loops were all shown to be within the green rating when applied to the EAST networks.

In this case the tenet analysis suggests structures in the system afford the vertical flow of information through the system. This indicates that enough feedback between agents is available to pass key information. One critique of the structure is the reliance on a single agent from the social network (the pilot) to pass critical information.

5.1.3 Feedback of current performance for future protections

The analysis suggests that the system is functionally dependent on knowing risks. Accordingly, the EAST networks indicate this is achieved by reporting of incidents and understanding hazards. The responsibility for reporting of incidents belongs to one agent, the pilot, who is positioned as the key agent in the social network. If reporting incidents is not fulfilled, the burden of accountability may fall towards the pilot as reporting incidents is exclusively their responsibility.

The tenet analysis indicates that more agents are required to protect against decremental effects. In this situation Decrementalism may lead to incidents occurring that the system ought to be aware of but are not due to an accepted degradation of information pathways. This may include non-reporting of recurring incidents, use of old risk assessment documentation or the erosion of the task itself where information is insufficient to learn how or what risks are present. In this case, non-human agents that can report feedback on performance (e.g. aircraft sensors, flight computers, video recording devices) to other agents in the system in addition to the pilot feedback may act as a safety measure against unsafe Decrementalism concerning incident reporting.

5.1.4 Accepting or acceptable: sensitive dependence of the system

Lastly the whole system is Sensitively Dependent on the Initial Condition to fly a Hawk Jet without a RAD ALT. The prevention of possible risks associated with flying too low can be maintained by providing equipment that provides an exact reading of height above sea in the cockpit. This draws attention to one way a system property can have a "cascading failure" effect when it interacts with other system properties in undesirable ways.

Currently the Hawk does not contain a RAD ALT as standard equipment. This may simply be a condition related to smaller aircraft where pilots support to land the plane with such equipment is deemed unnecessary for normal flight. The missile simulation certainly cannot be regarded as "normal flight" and surely then exceeds normal risks. Conversely, the costs of installing equipment such as a RAD ALT may not be deemed as financially practicable for the aircraft, when such flights are not within its normal operation.

5.2 Benefits for AI: representing safety boundaries as proactive leading indicators

The systems thinking tenets provide a valuable toolset given they expose and explain the theory behind accident aetiology. Conversely, they also assess and explain the key principles of safe systems. Given that Human Factors and Ergonomics is turning seriously toward proactive system performance, the systems thinking tenets provide a theoretical basis to do so.

One key finding is that the systems thinking tenets can be applied to assess how systems change over time. Indeed, this may be applied dynamically as a process of identifying leading indicators for industries that operate within safety critical domains. This is based on a measure of where

systems reside within acceptable performance boundaries. An increase in unsafe tenet properties may indicate proximity or indeed the movement toward the boundary of acceptable performance. Representing unsafe tenets may provide a method to understand where systems exist within acceptable performance boundaries. When considering the technologies such as AI, this may provide a timely measure of system performance prior to the experience of adverse events.

The capacity for AI to learn lessons based on the performance of the system is possible. Rasmussen's (1997) Risk Management Framework describes systems behaviours whereby the closer the system is to the safety boundary, there is an increase in unwanted system behaviours. A key challenge for Human Factors and ergonomics has been representing where a system resides within the safety boundary and the potential 'closeness' a system may be to experience adverse effects. The systems thinking tenets may provide a means to measure changes in systems behaviour, providing that the system can be modelled and interpreted at a level of value. This offers a potential representation of how gravitation towards the boundary of unacceptable performance occurs. Where there are more tenets present in an unsafe state, a system may be much closer to the safety boundary and undesirable variability. Further studies would be required to assess the tenets in several complex work systems overtime to determine the validity of this approach.

An example is provided below, which illustrates Rasmussen's (1997) safety boundary as a 'bubble' (see Figure 3). Within the bubble red, amber and green 'spheres' represent the performance indicators of the system as tenets. When indicators are closer to the boundary of defined work practices, they represent those that are in danger of crossing over performance boundaries resulting in adverse events. In this way the tenets may be mapped overtime to understand and track system performance and more importantly, intervene when the system is too close to the safety boundary.

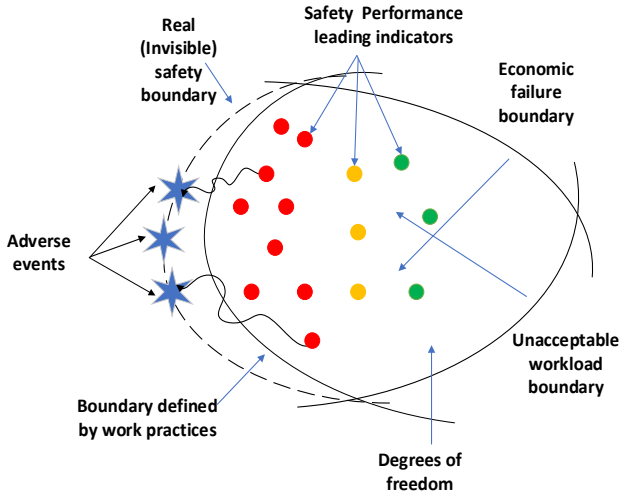

Figure 3. Representing proximity to performance boundaries using systems leading indicators (adapted from Rasmussen 1997).

5.3 AI and systems thinking: leading indicators and learning opportunities

An issue for the Hawk missile simulation system was regulating a height that was tolerable (for safety management) but also met the requirements of the training exercise for frigate crew.

Previously recorded incidents and near misses have been used to set safe flying altitudes by regulators of the system. Conversely, the frigate crew require the Hawk flight simulation to represent the authenticity of missile detection by flying as a low as possible. This poses a Constraint whereby the Hawk must fly as low as practicable to simulate a missile, but as high as needed to maintain safe altitude without a RAD ALT. This Constraint is currently maintained by regulators of the system where a strict flight parameter is upheld. However, the model of the system shows that this value may not change unless an incident occurs where the system is forced to review safe height above sea.

In contrast, AI may learn what safe and unsafe Constraints are by measuring and recording many metrics that have an interest in safe altitude in the absence of a RAD ALT. For example, the Constraint of safe height above sea (where it is as low as reasonably practicable) can be measured using information on, but not limited to, past incidents and near missies, typical or 'normal' flights, sea and weather conditions over time and pilot capabilities. The information can be used to learn what meets acceptable or tolerable performance maintaining the Constraint while also meeting the simulation requirements to detect a missile.

Where there are many variabilities that do not meet this, for example sea or weather conditions on a day of flight, pilot capabilities to navigate the conditions, AI may detect the impact of negative change and movement towards unsafe performance boundaries. This presents a means whereby AI can detect movement towards unsafe systems boundaries and time to review performance prior to incidents occurring.

AI may support safety critical systems through constant assessment of what does and does not meet performance boundaries by understanding how systems respond to safe and unsafe behaviours using the systems thinking tenets.

6 Conclusion

The paper has discussed how the core philosophies of accident causation may assess system performance proactively. The systems thinking tenets were applied to assess the Hawk missile simulation system modelled by EAST. As a set of leading indicators of system performance, the assessment showed the presence of more unsafe vulnerabilities than safe protections within the Hawk missile simulation system.

An additional benefit of the tenets explored in this paper, is that they may be applied to assess a systems proximity toward safety boundaries. When, more unsafe tenets exist in a system the gravitation toward safety boundaries, where accidents are more likely to occur, is more apparent. This is discussed as particularly useful for the introduction of AI where many unknowns exist, and where safety boundaries may be more disguised because of its inherent novelty. For the introduction of AI, the systems thinking tenets may be a useful proactive strategy to monitor performance but additionally allow AI systems to learn, adapt or respond to safety boundaries through monitoring

system performance prior to the occurrence of accidents in the system.

7 References

Carayon, P., Hancock, P., Leveson, N., Noy, I., Sznelwar, L., & Van Hootegem, G. (2015). Advancing a sociotechnical systems approach to workplace safety–developing the conceptual framework. Ergonomics, 58(4), 548-564.

Dekker, S., & Pitzer, C. (2016). Examining the asymptote in safety progress: a literature review. *International Journal of Occupational Safety and Ergonomics, 22*(1), 57-65.

Fogel, D. B. (2006). *Evolutionary computation: toward a new philosophy of machine intelligence* (Vol. 1): John Wiley & Sons.Richards, T.J. (1989): *Clausal form logic: an introduction to the logic of computer reasoning*. Sydney, Addison Wesley.

Grant, E., Salmon, P. M., Stevens, N. J., (2017) If Nostradamus were an Ergonomist: a review of ergonomics methods for their ability to predict accidents. In *Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2017 Annual Conference*. ISSN 2333-4959 (online). Available from http://hfes-europe.org

Grant, E., Salmon, P. M., Stevens, N. J., Goode, N., & Read, G. J. (2018a). Back to the future: What do accident causation models tell us about accident prediction? *Safety Science, 104*, 99-109.

Grant, E., Salmon, P. M., & Stevens, N. J. (2018b). The usual suspects? A novel extension to AcciMap using accident causation model tenets. *Theoretical Issues in Ergonomics Science*, 1-19.

Hancock, P. (2018). Some pitfalls in the promises of automated and autonomous vehicles. *Ergonomics*, 1-17.

Hancock, P. A., & Chignell, M. H. (2018). On human factors. In *Global perspectives on the ecology of human-machine systems* (pp. 14-53). CRC Press.

Hollnagel, E. (2012). *FRAM: the functional resonance analysis method: modelling complex socio-technical systems*: Ashgate Publishing, Ltd.

Kakimoto, T., Kamei, Y., Ohira, M., & Matsumoto, K. (2006). *Social network analysis on communications for knowledge collaboration in oss communities.* Paper presented at the Proceedings of the International Workshop on Supporting Knowledge Collaboration in Software Development (KCSD'06).

Leveson, N. (2004). A new accident model for engineering safer systems. *Safety Science, 42*(4), 237-270

Leveson, N. G. (2012). Complexity and safety. In *Complex Systems Design & Management* (pp. 27-39). Springer, Berlin, Heidelberg.

Leveson, N. (2013). The Drawbacks in Using the Term 'System of Systems'. *Biomedical Instrumentation & Technology, 47*(2).

Leveson, N., & Dekker, S. (2014). Get to the root of accidents: systems thinking can provide insights on underlying issues not just their symptoms. *Chemical Processing. PuttmanMedia*, 18-28

Maurino, D. E., Reason, J., Johnston, N., & Lee, R. B. (2017). *Beyond aviation human factors: Safety in high technology systems*: Routledge.

Mayo, R. C., & Leung, J. (2018). Artificial intelligence and deep learning–Radiology's next frontier? *Clinical imaging, 49*, 87-88.

Pelton J.N. (2019) Where Is Technology Leading Us: The Good, the Bad, and the Ugly. In: Preparing for the Next Cyber Revolution. Springer, Cham

Perrow, C. (2011). *Normal accidents: Living with high risk technologies*: Princeton University Press.

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). *Voice Interfaces in Everyday Life*. Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada.

Ramthun, A. J., & Matkin, G. S. (2014). Leading dangerously: A case study of military teams and shared leadership in dangerous environments. *Journal of Leadership & Organizational Studies, 21*(3), 244-256.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. *Safety Science, 27*(2), 183-213.

Rice, D. (2019). The Driverless Car and the Legal System: Hopes and Fears as the Courts, Regulatory Agencies, Waymo, Tesla, and Uber Deal with this Exciting and Terrifying New Technology. *Journal of Strategic Innovation and Sustainability, 14*(1).

Salmon, P. M., Read, G. J., & Stevens, N. J. (2016). Who is in control of road safety? A STAMP control structure analysis of the road transport system in Queensland, Australia. *Accident Analysis & Prevention, 96*, 140-151.

Salmon, P. M., Walker, G. H., M. Read, G. J., Goode, N., & Stanton, N. A. (2017). Fitting methods to paradigms: are ergonomics methods fit for systems thinking? *Ergonomics, 60*(2), 194-205. doi:10.1080/00140139.2015.1103385

Singh, S., Okun, A., & Jackson, A. (2017). Artificial intelligence: Learning to play Go from scratch. *Nature, 550*(7676), 336.

Stanton, N. A. (2008). Modelling command and control: event analysis of systemic teamwork. In C. Baber & D. Harris (Eds.). Aldershot, Hampshire, England; Burlington VT: Ashgate.

Stanton, N. A., Rafferty, L. A., & Blane, A. (2012). Human factors analysis of accidents in system of systems. *Journal of Battlefield Technology, 15*(2), 23. doi:http://www.oric.org.au/Research_Projects/MUARC_Accident%20Causation%20Analysis%20PPT.pdf

Stanton, N. A., & Walker, G. H. (2013). *Human factors methods: a practical guide for engineering and design*: Ashgate Publishing, Ltd.

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. doi:10.1080/00140139.2016.1232841

Stanton, N. A., Harvey, C., & Allison, C. K. (2019). Systems Theoretic Accident Model and Process (STAMP) applied to a Royal Navy Hawk jet missile simulation exercise. *Safety Science, 113*, 461-471.

Trist, E. (1981). The evolution of socio-technical systems. *Occasional paper, 2*, 1981.

Walker, G. H., Stanton, N. A., Baber, C., Wells, L., Gibson, H., Salmon, P., & Jenkins, D. (2010). From ethnography to the EAST method: A tractable approach for representing distributed cognition in Air Traffic Control. *Ergonomics, 53*(2), 184-197.

Walker, G. H., Stanton, N. A., Salmon, P. M., & Jenkins, D. P. (2008). A review of sociotechnical systems theory: a

classic concept for new command and control paradigms. *Theoretical Issues in Ergonomics Science, 9*(6), 479-499.
Waterson, P., Jenkins, D. P., Salmon, P. M., & Underwood, P. (2017). 'Remixing Rasmussen': The evolution of Accimaps within systemic accident analysis. *Applied Ergonomics, 59, Part B,* 483-503. doi:https://doi.org/10.1016/j.apergo.2016.09.004
Underwood, P., & Waterson, P. (2013). Systemic accident analysis: Examining the gap between research and practice. *Accident Analysis & Prevention, 55*(0), 154-164. doi:http://dx.doi.org/10.1016/j.aap.2013.02.041
Woods, D. D., & Hollnagel, E. (2017). Prologue: resilience engineering concepts *Resilience engineering* (pp. 13-18): CRC Press.