
Article:
Burton, Simon, McDermid, John Alexander orcid.org/0000-0003-4745-4272, Garnett, Philip orcid.org/0000-0001-6651-0220 et al. (1 more author) (2021) Safety, Complexity, and Automated Driving:Holistic Perspectives on Safety Assurance. Computer. pp. 22-32. ISSN: 0018-9162

https://doi.org/10.1109/MC.2021.3073430

Safety, complexity and automated driving – holistic perspectives on safety assurance

Simon Burton
Fraunhofer Institute for Cognitive Systems, Munich, Germany
University of York, York, UK

John McDermid
University of York, York, UK

Philip Garnet
University of York, York, UK

Rob Weaver
Independent Consultant, Canberra, Australia

***Abstract*—This paper extends safety assurance approaches for automated driving by explicitly acknowledging the complexity associated with the emergent system behaviour and its wider socio-technical context. We introduce a framework for reasoning about factors that contribute to this complexity as a means of more effectively structuring the discussion and thus aligning the inter-disciplinary perspectives required to achieve a socially and legally acceptable level of residual risk. The framework is illustrated by performing a post-hoc analysis of a high profile accident involving a prototypical automated driving vehicle. We then apply the framework to analyse the potential risks associated with the public introduction of Automated Lane Keeping Systems (ALKS) for which regulation is currently being developed to prepare for deployment of the systems on public roads. This analysis leads of specific recommendations both for the ALKS in particular and safety assurance methodologies for automated driving systems in general.**

HUMAN ERROR is by far the greatest contributing factor to fatal accidents on U.K. roads [1], whilst environmental effects (8%) and vehicle defects (2%) play a relatively insignificant role in comparison. Automated driving systems (ADS) have the potential for making roads significantly safer by optimizing traffic flow, recognizing and reacting to hazards on the route ahead and limiting the impact of inattentive and unreliable human drivers. There is currently a drive to introduce Automated Lane Keeping Systems (ALKS), with a public consultation ongoing in the UK [2] at the time of writing.

This paper extends the discussion on the safety of automated driving systems beyond the traditional focus of the engineering community. We explicitly do not restrict our analysis to technological aspects but consider the role of governance, management and operation as well the role of human factors in order to establish a holistic

 Published by the IEEE Computer Society

view of safety. In doing so, we acknowledge the complexity of the problem at hand and the fact that a "vehicle-centric" focus to engineering safe automated driving may be inadequate.

As part of a study [3] commissioned by Engineering X, an initiative coordinated by the Royal Academy of Engineering and supported by the Lloyd's Register Foundation, the authors were tasked with producing a framework (hereafter referred to in this paper as the *safer complex systems framework*) to provide conceptual clarity around the factors that lead to systemic failures in complex systems which have a safety impact.

This paper builds on the results of the study and is structured as follows: in the following section we discuss the relation of complexity to the safety of autonomous systems. We then introduce a framework for identifying factors that impact the safety of complex systems. The framework is illustrated by using it to model an incident involving a prototypical automated vehicle and then used to analyse the potential risks associated with interactions between systems when deploying ALKS on public roads. A set of recommendations for addressing complexity in the safety assurance of automated driving is provided as is an agenda for future work.

COMPLEXITY AND SAFETY

What is a complex system?

Complex systems theory defines a system as *complex* if some of the behaviours of the system are *emergent* properties of the interactions between the parts of the system, where the behaviours would not be predicted based on knowledge of the parts and their interactions alone.

From the perspective of complexity science there are a number of characteristics that are shared by most, if not all, complex systems. These are variously described and defined in [4, 5] and include, amongst others:

- **Semi-permeable boundaries:** The boundaries between the system and the environment are dependent on the scope of the system under consideration, known as the *system of interest*. This may vary depending on the objectives of the analysis. For example, if focusing on the functional performance of an automated vehicle the system could be viewed as a set of electronic components which sense the environment, decide on control actions and implement them via actuators. However, when considering a mobility service as a whole, the system includes other traffic participants, emergency services and city or highway infrastructure as well as the vehicle and impact on the use of public transport [6].
- **Non-linearity, mode transitions and tipping points:** The system may respond in different ways to similar input depending on its state or context. Non-linear behaviour can also be caused by *coupled feedback* both within the system of interest and between the system and its environment. It is common to talk about complex systems going through critical transitions widely referred to as *tipping points*. Tipping points can also be transitions into unsafe states, and these can be emergent properties of the system itself. The seemingly spontaneous occurrence of traffic jams and stop-start traffic on motorways are examples of such behaviour within traffic systems.
- **Self-organisation and *ad-hoc* systems:** Systems can also emerge in an *ad-hoc* manner, through a convergence of parts perhaps by a process of self-organisation, or self assembly. Here the (semi-permeable) boundary may change as the system evolves. The adaption in behaviour of human road users in response to automated vehicles is an example of self-organisation, where the humans become part of a larger *ad-hoc* system. The ability of traffic to spontaneously respond to approaching emergency vehicles, even at complex intersections, is an example of *ad-hoc* self-organisation.

Safety of complex systems

Traditional systems safety engineering focuses on component faults and their interactions with other system components and therefore requires some model of the system in order that these interactions can be analysed. In contrast, complex systems can give rise to systemic failures which do not necessarily arise from faults in individual system parts. This bears a strong and deliberate relationship to the definition of complex systems and the notion of emergence. Systemic failures originate from the interactions between the parts of the system (their behaviours)

and interaction with or dependencies on the environment, rather than faulty components, buggy software functions or wear and tear – although such things might be contributing, though not sufficient, causes for a systemic failure.

The difficulty of arguing the safety of an ADS lies in the inherent complexity and unpredictability of the ever-changing environment in which it operates. To compound the problem, the system observes this complex, unpredictable environment using sensors that themselves have inherent inaccuracies due to the physical limitations of their sensing modalities. This uncertainty is countered by using multiple sensor types and algorithms that make use of heuristics or ML to interpret the sensing data. However, these algorithms are themselves inherently imprecise and introduce an additional level of uncertainty [7]. The unpredictable nature of the impact of the vehicle's actions on its environment (e.g. the reactions of other drivers and road users) "closes the loop" to the complex environment to be interpreted by the vehicle. Thus implementing ADS brings with it potential for systemic failures due to interactions between these uncertainties within the perception and control cycle.

Risk reduction measures, or controls, to reduce the probability of safety-related failures can include engineering changes at design time, or procedures and processes implemented during operation. Controls can be grouped in very broad terms into those that enable:

- **Robustness** – the ability of a system to cope with foreseen events.

which we contrast with:

- **Resilience** – the ability of a system to absorb the unforeseeable and remain unchanged.

Both resilience and robustness are tools for reducing risk, with resilience more important in dealing with the uncertainties arising from complexity. Complexity science uses these terms rather differently and, for example, resilience is used to mean that the system returns to its original state or maintains its original function. Here, resilience might mean that the system changes behaviour (or even purpose) but continues to operate safely in the presence of unforeseen events.

THE SAFER COMPLEX SYSTEMS FRAMEWORK

The framework we propose provides a structure for reasoning about factors that contribute to systemic failures due to complexity and contextualises measures and controls to manage risk.

As visualised in Figure 1, the central axis of the *safer complex systems framework* shows a flow from **causes** of system complexity via their **consequences** to **systemic failures**. This is analogous to the progression from faults to erroneous system states to system failures underlying traditional functional safety engineering as promoted by Laprie et al [8]. However, as noted above, systemic failures arise out of emergent properties of the system caused by complexity, not from faults in individual system elements, and that the inter-dependencies between system elements as well as the causes and consequences of complexity are more subtle than a simple cause-effect relationship – however, the visualisation of Figure 1 is chosen for ease of explanation and discussion.

The emergence of systemic failures can be tempered by **controls** at **design-time** and during **operation**. These reduce the likelihood that systemic failures arise by either suppressing the causes of complexity or by reducing the likelihood that emergent complexity leads to the failure to maintain a system objective. The framework also recognises **exacerbating factors** that can make systemic failure more likely by either amplifying the consequences of system complexity or undermining control measures. Inherent uncertainty in our knowledge of the system and its boundaries, the models we use to reason about the systems, or the technology itself (e.g. use of machine learning techniques) are examples of exacerbating factors that can increase the consequences of complexity or undermine controls.

In many cases, causes of complexity as well as the controls for managing safety are regulatory, organisational or financial, instead of, or in addition to, technical. Furthermore, not all systems are explicitly engineered; they can also arise from *ad hoc* interactions between systems or components previously considered unrelated. This requires radically different approaches and viewpoints to previously applied safety engineering and management techniques that were based on clearly

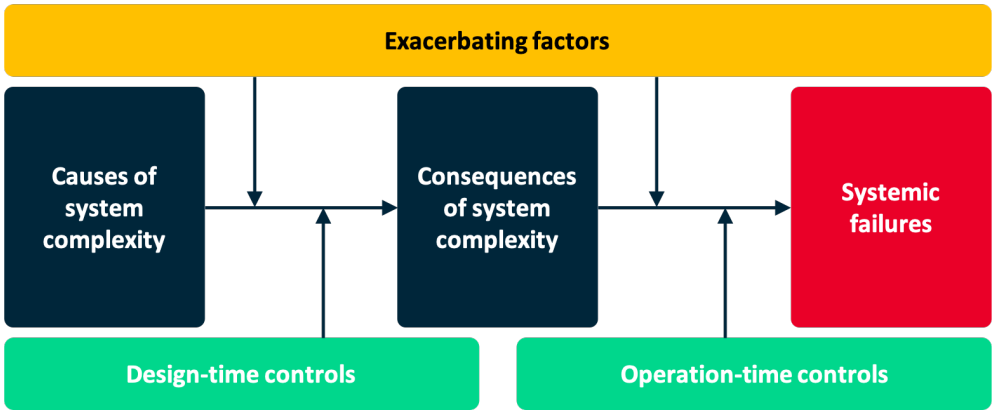

Figure 1. The Safer Complex Systems Framework

defined system boundaries. The framework is intended to address both designed and *ad hoc* systems and considers a system through the lens of the following strongly interacting viewpoints. These can be seen as layers within the overall model, akin to those found in Rasmussen's risk management framework [9].

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

Whilst developing the framework, examples from a number of domains including aerospace, mobility, healthcare and supply networks [3] were analysed to identify common categories of causes, consequences, systemic failures, exacerbating factors and controls across these three layers. This resulted in a set of guide-words that could be used as part of a structured analysis performed by interdisciplinary experts and can also be based on a specific investigation of previous system failures. The framework is not intended to replace existing safety analysis and management approaches. Instead, it provides an additional perspective to allow the perceived system boundaries, stakeholders and influencing factors to be called into question, thereby providing a more robust basis for finding gaps in current safety thinking and providing context for more specific safety analyses.

APPLICATION OF THE FRAMEWORK

We illustrate the framework by examining the issues surrounding the introduction of ADS. For the purposes of this paper we understand an ADS as a system that takes over control of the vehicle while driving under a given set of conditions. During this time, the driver can direct their attention to other pursuits while the system takes control of the vehicle. The driver must be available to take over control when the boundary

of the Operational Design Domain (ODD) is met.

An essential first step in the analysis is to determine an initial scope of the system of interest. Note that, as a result of the analysis, factors outside the assumed system scope could be determined to be relevant, leading to a revision of that scope as part of an iterative process. For the purposes of the example here the system scope shall be defined as follows:

- **System scope:** Traffic infrastructure, traffic participants, emergency services, regulations and responsible authorities, including both manually driven and automatically controlled vehicles.

The next step in the application of the framework is to analyse factors that lead to (intractable) complexity and therefore the potential for systemic failures. We illustrate the framework here by performing a *post-hoc* analysis of an accident to better understand the relationship between the causes that lead to the system failure. The description of the accident summarised below is based on the US National Transportation Safety Board accident report and recommendations [10].

In March 2018, an automated test vehicle operated by Uber Advanced Technologies Group (Uber ATG) was involved in an accident in Tempe, Arizona, that fatally injured a pedestrian who was crossing a dual carriageway while pushing a bicycle. The circumstances surrounding this accident highlight many of the risks involved in introducing automated driving technologies as well as the potential for systemic failures at the task and technical, management and governance layers.

Analysis of vehicle data demonstrated that up until the impact, the vehicle variously misclassified the pedestrian as a vehicle, unknown object and bicycle. On each new classification, the object trajectory prediction algorithm would reset and assign a new classification-dependent trajectory prediction.

1.2 seconds before impact, the system identified an unavoidable collision. However, in order to avoid the consequences of false-positive mis-classifications, the system was designed to suppress any braking manoeuvres in such a case, due to the assumption that an attentive operator would take control. The safety driver was, at the time, viewing content on her mobile phone and was therefore not able to react to prevent the impact. Furthermore, emergency braking systems pre-installed within the vehicle had been deactivated in order not to conflict with the prototypical functions under test.

The accident report [10] identifies inattentiveness of the operator as the most probable cause of the crash. However, it also identified a number of additional contributing factors, including inadequate safety risk assessment procedures at Uber ATG, and ineffective oversight of the vehicle operators, including lack of mechanisms for addressing operators' automation complacency. Additional factors were identified as the ambiguous nature of the piece of ground separating the directions of the carriageway which appeared to include pedestrian walkways, and ineffective oversight of automated vehicle testing by Arizona's Department of Transportation.

The *safer complex systems framework* can now be used to identify causes, consequences and exacerbating factors leading to the accident. The results of this analysis is summarised in Figure 2. The following manifestations of system complexity were identified:

- **Governance:** Rapid technological change, insufficient competence and awareness regarding associated risks and the competing objectives of accommodating business needs vs. regulatory responsibilities led to a loss of regulatory control at the state-level and inappropriate deployment decisions. This ultimately led to an increased risk to other traffic participants and an avoidable accident.
- **Management and operation:** Inadequate engineering and release processes related to the novelty and complexity of the safety issues involved, coupled with market pressure, transference of responsibility to an inadequately trained and supervised operator led to not only a technically inadequate system but also operational procedures that did not adequately account for (potentially unanticipated) classes of risk.
- **Task and technical:** The complexity of the environment and behaviour of different agents within it was underestimated and emergent behaviours related to the interaction of the system

and (the attentiveness of) the safety-driver as well as of pedestrians and their surroundings were not adequately considered leading to a failure of the core driving function as well as the primary backup which, in this case, was the safety driver.

From this perspective, functional insufficiencies of the system at the technical level as well as the behaviour of the safety driver can be seen as emergent properties. They arose, at least in part, from the management and governance levels and the apparent failure of the duty holders to understand and manage the risks associated with operating such systems. There were insufficient measures in place at the governance and management level to constrain the emergent risk of deploying the technical system in its environment. This may, in part, be due to a lack of understanding (competency gap) of the system scope to be considered as well as the potential for emergent behaviours within the system that included the vehicle, driver, pedestrian and road layout as interacting constituent parts. The example also demonstrates the conflicting pressures to promote innovation in technologies such as automated driving that have the potential for improving overall road safety, while in parallel managing the risk of integrating such technologies into existing traffic systems with an insufficient understanding of emergent behaviours.

Consequences for the deployment of ALKS on public roads

In this Section we apply the *safer complex systems framework* to a discussion of the risks associated with the deployment of ALKS [2] systems onto public roads in the U.K. In doing so, we describe the predicted effectiveness or otherwise of existing control measures and identify where further measures are required. The analysis was based on a review of regulations [2], standards [11], industry best practices [12] and lessons learned from previous accidents such as [10]. The set of guide-words identified when developing the framework were applied to identify risk factors and allocate them within the framework to better understand their impact and inter-relationships. The analysis resulted in the identification of a number of themes where distinct relationships were found within the various components of the framework. These included the difficulty in defining a tolerable level of residual risk and liability for the systems, the issue of calibrated trust [13] and automation complacency, an appropriate definition of the operating domain, risk transference between different stakeholders and interaction with existing traffic systems.

Figure 3 shows *causes* of (unsafe) complexity arising from multiple jurisdictions, semi-permeable system boundaries, heterogeneity and inter-connectivity on the safety of ALKS. At the *governance* level *multiple jurisdictions* refers to the regulations for smart motorways and ALKS itself. At its simplest, the smart motorways rules mean that vehicles should not travel in lanes when a red X is shown and to move into other lanes; the ALKS regulations do not address such signs and do not require vehicles to be able to change lanes. A *systemic failure* that can be linked directly to this is lack of clear allocation of liability.

A further example of problems is the likely heterogeneity and inter-connectivity between vehicles fitted with ALKS giving rise to emergent properties so that behaviour so the system-of-systems is no longer predictable – particularly if we consider how manually driven vehicles might behave if they see an ALKS-fitted vehicle proceed past a X. A *design-time control* to address this and the above governance-level issue involves refining the ALKS specification to deal with smart motorway infrastructure, but also to ensure sufficient consistency between different manufacturers' vehicles that they operate safely in a system-of-systems. Although shown as a *task & technical* level control this may also need to be reflected at the *management* and *governance* level, e.g. in regulations, if it is to be effective.

A related *operation-time control* is to enable the infrastructure to "orchestrate" the behaviour of the ALKS-equipped vehicles, for example to ensure they all move in the same direction when approaching a lane with a X. Such "orchestration" could also deal with interaction between ALKS and emergency services by, for example, forcing the ALKS-equipped vehicles to create an "extra lane" by moving in opposing directions and thus allowing emergency vehicles through.

Thus, greater inter-connectivity is likely to be needed to enable safe introduction of ALKS.

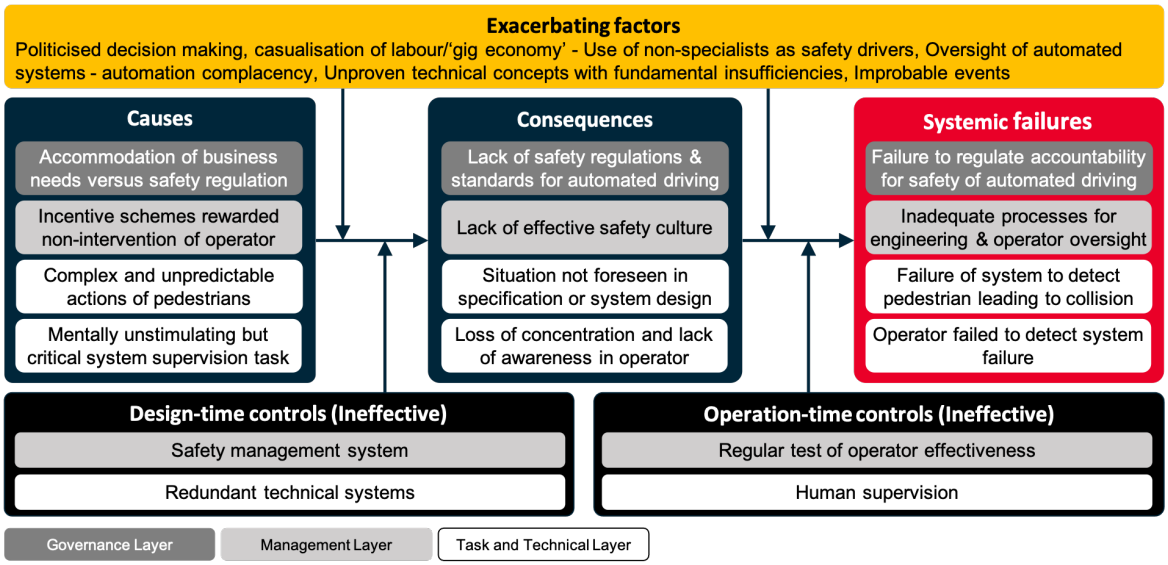

Figure 2. Freak accident or causal inevitability? The Uber Tempe fatality

However, this will inevitably lead to additional emergent properties including issues of cyber-security and a consequential interplay between safety and security. To address such changes requires further iteration, thinking through the impact of the changes, including the possibility of cyber-security weaknesses introducing common-mode failures.

Due to the rapid technological changes driving the transformation of the mobility sector, it is not feasible to expect that traditional approaches to standards development will keep pace with the rate of change. Therefore, outcome-based regulation that stipulates requirements on *what* to argue instead of *how* to argue safety is to be developed; it should take a systems-oriented view with additional focus on arguing the effectiveness of controls for reducing risk due to system complexity. Published standards and regulations should be supported by publicly available specifications that provide more specific guidance and document current industry consensus on topics such as assurance activities for machine learning in an automated driving context. These specifications can be developed in a more agile manner than full standards and can therefore be continuously updated to reflect the state-of-the-art.

A consensus on safety targets for automated driving must be actively developed including a diverse range of stakeholder perspectives beyond just manufacturers and technical approval authorities. This should consider both quantitative targets (e.g. based on accident statistics) as well as qualitative measures (based on engineering practices and operation-time controls) for achieving acceptable levels of residual risk. This will require cross-disciplinary dialogue involving not only technical but also legal and ethics experts [14]. Wider engagement with the public in general is also required in order to consider the perspectives of those most impacted by risk, and also to gain an understanding of the expectations and assumptions made on the systems by the users. This is required in order to reach a level of trust and acceptance of the systems, without which the safety benefits of increased automation will also not be realised.

Consequences for safety assurance of automated driving

The manifestations of complexity described in previous sections introduce uncertainty across the entire assurance process; models of the ODD used to design and validate the system are inevitably incomplete and imprecise due to the complexity of the environment; technology used within the systems both in terms of the sensors and actuators as well as the algorithms themselves are inherently imprecise (e.g. based on the use of Machine Learning). Furthermore, due to a lack of

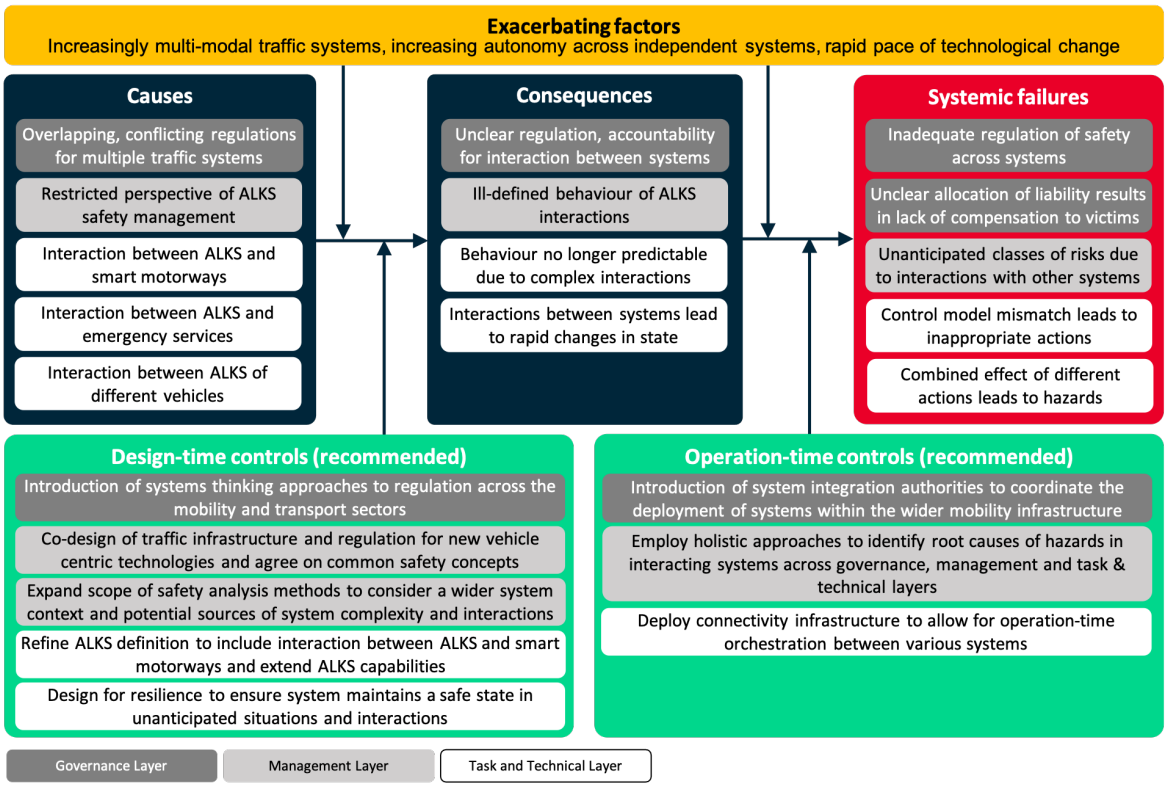

Figure 3. Analysis of interactions between ALKS and other traffic systems

clear definitions of safety targets and the lack of established best practice, the assurance case itself may include assurance gaps, leading to inconclusive arguments. Complexity and the resulting uncertainty must therefore be considered within all phases of the safety management processes and means must be developed for arguing that the residual risk of such systems is nevertheless tolerable.

During *domain analysis*, an understanding is developed of the safety-relevant properties that must be maintained within the chosen operating environment. In addition to properties of the ODD itself, critical for ensuring the performance of perception and control functions, this phase must also ensure a sufficient understanding of societal and legal expectations on the system for it to be considered safe enough. This analysis must take place within an assumed scope of the system under consideration and its interactions with its environment. This system scope shall be continuously validated and adjusted to ensure that critical interactions are considered in the safety assurance approach.

The *system design* refines the expectations on the system discovered during the domain analysis into technical requirements on the system and identifies a system design capable of supporting the system's safety goals. The *safer complex systems framework* can provide indicators related to risk associated with the complexity of the system and its context, its operation and interaction with human actors that should be considered during detailed safety analyses.

Causal approaches to safety analysis such as Fault Tree Analysis (FTA) [15] are based on a model [8] of how faults in individual system components cause an erroneous system state (error) that may subsequently lead to a failure of the system's service as perceived by its users. However, one of the consequences of complexity is the presence of unknown and unknowable faults and causes of systemic failures as well as a high level of inter-connectivity and non-linear interactions. There is also a need to be able to model the impact of uncertainties both in the environment as well as the internal behaviour of the system [16]. System-theoretic approaches

such those recommended by Rasmussen [9], Leveson's System Theoretic Accident Methods and Processes/System Theoretic Process Analysis (STAMP/STPA) [17] and Hollnagel's Functional Resonance Analysis Method (FRAM)[18] take a more holistic view of risk factors within a system across technical, management and governance layers. These methods have the potential to better consider the causes and consequences of complexity within the system across these layers. We therefore strongly recommend that these methods are applied when preparing for the deployment of ALKS onto public roads. Monkhouse et. al [19] recently proposed an enhanced vehicle control model that considers the shared cognitive nature of the driving task for driver assistance systems with limited autonomy. This approach allows the subtle interactions between the task (human-controlled activities) and technical (system-controlled activities) within our framework to be more systematically analysed, for example extending STAMP [17] or FRAM [18] type methods.

An analysis of complexity factors that could lead to systemic failures can provide information for *verification and validation* by determining sets of assumptions that must be confirmed and specific properties that must be validated during field tests and operation. Relating back to the Uber ATG accident described previously, this could include validation of assumptions made regarding the performance of the safety driver or about the behaviour and occurrence probabilities of pedestrians on certain types of roads.

Statistical arguments based on miles driven between incidents during field-based tests become both unfeasible and ineffective due to the effort required to collect the data and the difficulty in ensuring sufficient coverage of edge cases and critical situations. The increase in use of simulation during the design and validation of the systems allows for a more targeted testing of critical and rare situations. However, such approaches require additional arguments regarding the accuracy and the ability to extrapolate the results of simulation into the target domain.

We consider the greatest benefit of explicitly considering complexity factors lies in the formulation of the *assurance case* [20] and more specifically the reduction of assurance gaps. Application of the *safer complex systems framework* can lead to an assurance case that better reflects the actual system context and risks associated with the system and its operating context. This could be achieved by integrating consideration of complexity factors throughout the claims and evidence provided in the assurance or by formulating specific claims and targeted evidence focusing on the causes of systemic failures in the system.

Safety assurance of automotive systems currently places a strong focus on design-time controls and type approval. However, as the complexity and scope of the systems increases, and with it the sensitivity to an ever evolving environment, it is unrealistic to believe that an adequate level of safety can be achieved before the system is deployed that can be maintained over the vehicle's lifetime – without ongoing controls and the potential for updates to the system. Operation-time measures are required for ensuring the safety of the systems that includes the measurement of critical observation points within the system (leading indicators of systemic failures) as well as whether assumptions made regarding the ODD and therefore the validation approach continue to hold. The assurance case for the system should be continuously evaluated and refined, based on experiences in the field and changing expectations on the system. This holds true for automated driving applications but also to connected traffic infrastructure in general.

CONCLUSIONS

Assuring the safety of autonomous vehicles is a complex endeavour and the deployment of automated vehicles within a public traffic infrastructure must be recognised as a complex system requiring complexity thinking. By this we not only mean that it is just technically difficult, or involves many resource intensive tasks that must somehow be managed within feasible economic constraints. Both are true. But autonomous vehicles and their wider socio-technical context demonstrate characteristics of complex systems in the stricter sense of the term. This has a huge impact on our ability to argue the safety of such systems.

This paper proposed a framework by which factors impacting the complexity of the system, thus leading to systemic safety failures, can be

identified and used to inform a safety assurance process. We conclude from the analyses described in this paper that ensuring and demonstrating the safety of automated driving systems requires a more comprehensive and holistic view of safety than for previous generations of vehicle electronic control systems. A systems-oriented approach that acknowledges complexity and includes coordinated measures across governance, management and task and technical layers is required in order to reach an adequate level of safety for automated driving systems. This will require closer collaboration between domains such as automotive manufactures and suppliers, communication and highway or city infrastructure as well as a better understanding of dependencies across the three layers and stakeholders within the framework which includes the role of the general public.

The *safer complex systems framework*, at this stage in its development, seeks to provide an accessible overview of the *factors* that influence the safety of complex systems. As presented, the framework indicates only the highest-level dependencies between elements of the framework. Further work will involve enriching the framework by integrating various safety analysis methods as well as domain-specific risk models in order to allow the framework to be integrated into safety analysis and management during system design. Work is also ongoing to validate the framework in other domains including urban air mobility and healthcare. Most significantly though, the authors see the strongest need in establishing a systems-thinking mindset that acknowledges complexity and uncertainty both at the governance and management layers as ultimately it is here where the levers are most effective to ensure that our traffic systems remain safe, and become even safer through the introduction of autonomous technologies.

ACKNOWLEDGMENT

This project was supported by the Royal Academy of Engineering and the Lloyd's Register Foundation as part of the EngineeringX programme. The authors would like to thank all those who contributed to the project.

References

[1] *U.K. Department for Transport, Contributory factors for reported road accidents (RAS50)*. https : https : / / www . gov . uk / government / statistical - data - sets / ras50 - contributory - factors#contributory - factors - for-reported-road-accidents-ras50---excel-data-tables. 2020.

[2] UK Department for Transport and Centre for Connected Autonomous Vehicles. *Safe use of Automated Lane Keeping System on GB motorways: call for evidence*. https:// www . gov . uk / government / consultations / safe - use - of - automated - lane - keeping - system - on - gb - motorways - call - for - evidence. Accessed: 2020-10-18.

[3] Simon Burton et al. *Safer Complex Systems – An Initial Framework*. https : / / www . raeng . org . uk / global / international - partnerships/engineering-x/safer-complex-systems. Royal Society of Engineering, to be published in January 2021.

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

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

[6] Christina Pakusch et al. "Unintended effects of autonomous driving: A study on mobility preferences in the future". In: *Sustainability* 10.7 (2018), p. 2404.

[7] Simon Burton, Lydia Gauerhof, and Christian Heinzemann. "Making the Case for Safety of Machine Learning in Highly Automated Driving". In: *Computer Safety, Reliability, and Security*. Ed. by Stefano Tonetta, Erwin Schoitsch, and Friedemann Bitsch. Cham: Springer International Publishing, 2017, pp. 5–16. ISBN: 978-3-319-66284-8.

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

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

[10] National Transportation Safety Board. “Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian Tempe, Arizona March 18, 2018”. In: (2019).

[11] International Organization for Standardization. *Road vehicles — Safety of the intended functionality*. Tech. rep. ISO/PAS 21448:2019. ISO, 2019.

[12] International Organization for Standardization. *Safety and cybersecurity for automated driving systems — Design, verification and validation*. Tech. rep. ISO/PRF TR 4804:2020. ISO, 2020.

[13] Liza Dixon. “Autonowashing: The Greenwashing of Vehicle Automation”. In: *Transportation Research Interdisciplinary Perspectives* 5 (2020), p. 100113.

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

[15] Glenn Bruns and Stuart Anderson. “Validating safety models with fault trees”. In: *SAFECOMP’93*. Springer, 1993, pp. 21–30.

[16] Roman Gansch and Ahmed Adee. “System Theoretic View on Uncertainties”. In: *DATE 2020*. 2020.

[17] Nancy Leveson. *Engineering a safer world: Systems thinking applied to safety*. MIT press, 2011.

[18] Erik Hollnagel. *FRAM, the functional resonance analysis method: modelling complex socio-technical systems*. Ashgate Publishing, Ltd., 2012.

[19] Helen E Monkhouse, Ibrahim Habli, and John McDermid. “An Enhanced Vehicle Control Model for Assessing Highly Automated Driving Safety”. In: *Reliability Engineering & System Safety* (2020), p. 107061.

[20] International Organization for Standardization. *Systems and software engineering — Systems and software assurance*. Tech. rep. ISO/IEC/IEEE 15026:2019. ISO, 2019.