TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Engineering and Design

The Contribution of Automated Driving to Road Traffic Safety – A Resilience Engineering Approach

Niklas Grabbe, M.Sc.

Vollständiger Abdruck der von der TUM School of Engineering and Design

der Technischen Universität München

zu Erlangung eines

Doktors der Ingenieurwissenschaften (Dr.-Ing.) genehmigten Dissertation.

Vorsitz: Prof. Dr.-Ing. Birgit Vogel-Heuser

Prüfer der Dissertation:

1. Prof. Dr. phil. Klaus Bengler
2. Prof. Erik Hollnagel, Ph.D.
3. Prof. Dr. phil. habil. Oliver Sträter

Die Dissertation wurde am 07.09.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Engineering and Design am 30.01.2024 angenommen.

Acknowledgments

The last years were an intense but fantastic journey within research, teaching, and project work, within books, articles, models, simulations, people, conferences, and foreign cultures - full of learning on a scientific and personal level. At this point, I would like to reflect on the people supporting me, which I am so grateful to thank here.

First of all, I would like to thank my supervisor Professor Klaus Bengler for the opportunity to accomplish a dissertation on the complex interplay of road safety and automated driving at the Chair of Ergonomics. I very much appreciate the freedom to evolve my own research interests and to educate myself, including research, teaching, and networking. A special thanks goes to Professor Erik Hollnagel. It has been an honour to have you, as the FRAM inventor and one popular researcher in the field of human factors and safety, for being available as a second examiner of this thesis. Furthermore, I would like to thank my Mentor Christian Gold for his support and valuable advice, especially at the beginning of my journey.

I had great colleagues at the institute who also became friends, for which I am very grateful. You always ensured a friendly and helpful atmosphere inviting to come to work. In particular, I want to thank Burak Karakaya, Michael Rettenmaier, Lorenz Steckhan, Alexander Feierle, and Deike Albers for fruitful, scientific discussions, reading my papers or thesis enriched with valuable insights and comments, and help when and wherever needed.

I would also like to thank the community of FRAMily for stimulating discussions, above all, at the annual meeting in Malaga in 2019. Special thanks go to Riccardo Patriarca and Arie Adriaensen for deep and nerdy discussions concerning FRAM.

Moreover, I thank all the students, peers, and subjects collaborating or participating in my research.

Last but not least, a huge thanks goes to my family because my successes are definitely a bit of yours.

Abstract

Automating the driving process, leading to self-driving cars, has remained a prominent human aspiration. In fact, this goal is currently achievable: substantial advancements have been achieved in realising the necessary vehicle technologies promising an increase in traffic safety, efficiency, and driving comfort. However, introducing automation into other domains has shown that anticipated benefits may also be accompanied by unexpected, adverse effects, causing as many problems as solved. The evidence of the safety of highly automated driving is still lacking, a situation often termed the "approval trap", thus necessitating the development of novel testing approaches. These approaches must consider that automated vehicles face mixed traffic, representing at least a long transition phase up to autonomous driving, acknowledging the complex and intractable adaptations in the socio-technical system that traffic represents. Thus, a differentiated understanding of the patterns in road traffic leading to accident development and avoidance is required to prevent adverse automation surprises, which is currently lacking.

This thesis investigates the potential benefit of a resilience engineering approach and its reasoning for enhancing road safety in the context of automated vehicle introduction by taking a systems thinking mindset using the functional resonance analysis method (FRAM) and a safety-II perspective. First, the Sections 1-4 provide the foundational reasoning for that. The body mainly has a publication-based structure. A literature review combined with a case study, Article 1 (Section 5), identifies and methodologically evaluates FRAM as a suitable method for differentiating patterns to assess road safety related to human and automated driving. In Article 2 (Section 6), relevant test scenarios are deduced based on reasonable criteria for scenario selection in the safety assessment of automated driving in which FRAM should be applied. Article 3 (Section 7) uses FRAM in an overtaking scenario to create a model understanding the patterns of accident development and accident avoidance to give system design recommendations and essential insights for the validation process of automated vehicles. Article 4 (Section 8) evaluates the model and method in terms of validity to assess the credibility of the results and the applicability, respectively. In Section 9, the research findings of Articles 3 and 4 are revisited using a pure function-based validation approach differentiating instantiations to get an enhanced comprehension with regard to the FRAM model's credibility and to identify enhanced patterns that can be used for system improvement. Ultimately, Section 10 discusses the results concerning system design and validation, method evaluation, and industrial application.

The research findings demonstrate that resilience engineering, including FRAM and safety-II, is an invaluable and essential missing approach to assessing the safety of automated driving in road traffic, mainly by studying interactions in view of emergentisms. FRAM is a promising approach addressing the nitty-gritty - to identify the patterns that facilitate the system's adaptive capacity inevitable for safe and efficient performance in socio-technical systems. A solid foundation on a small-scale addressing the right problem is laid where enhancements and extensions have to be researched in the future of how to deploy the approach in the industry on a large-scale.

Zusammenfassung

Die Automatisierung des Fahrens, die zu selbstfahrenden Autos führt, ist nach wie vor ein wichtiges Bestreben der Menschheit. Tatsächlich ist dieses Ziel derzeit erreichbar: bei der Realisierung der erforderlichen Fahrzeugtechnologien, die eine Erhöhung der Verkehrssicherheit, der Effizienz und des Fahrkomforts versprechen, wurden bereits erhebliche Fortschritte erzielt. Die Einführung der Automatisierung in anderen Bereichen hat jedoch gezeigt, dass die erwarteten Vorteile auch von unerwarteten, negativen Auswirkungen begleitet sein können, die ebenso viele Probleme verursachen wie lösen. Die Sicherheit des hochautomatisierten Fahrens ist nach wie vor nicht nachgewiesen, eine Situation, die oft als "Zulassungsfalle" bezeichnet wird und die Entwicklung neuartiger Absicherungskonzepte erforderlich macht. Diese Ansätze müssen berücksichtigen, dass automatisierte Fahrzeuge mit gemischtem Verkehr konfrontiert sind, was zumindest eine lange Übergangsphase bis zum autonomen Fahren darstellt, wobei die komplexen und flüchtigen Anpassungen im soziotechnischen System, das der Verkehr darstellt, zu berücksichtigen sind. Daher ist ein differenziertes Verständnis der Muster im Straßenverkehr, die zur Unfallentstehung und -vermeidung führen, erforderlich, um negative Überraschungen durch die Automatisierung zu vermeiden, was derzeit fehlt.

Diese Arbeit untersucht den potenziellen Nutzen eines Resilience-Engineering-Ansatzes und dessen Schlussfolgerungen für die Erhöhung der Verkehrssicherheit im Zusammenhang mit der Einführung automatisierter Fahrzeuge, indem sie eine systemorientierte Denkweise unter Verwendung der funktionalen Resonanzanalysemethode (FRAM) und einer Safety-II-Perspektive einnimmt. Zunächst werden in den Abschnitten 1-4 die grundlegenden Schlüsse und Argumente zu diesem Thema dargelegt. Der Hauptteil ist hauptsächlich auf der Grundlage von Veröffentlichungen aufgebaut. In Artikel 1 (Abschnitt 5) wird anhand eines Literaturüberblicks und einer Fallstudie FRAM als geeignete Methode zur Unterscheidung von Mustern für die Bewertung der Verkehrssicherheit in Bezug auf menschliches und automatisiertes Fahren identifiziert und methodisch bewertet. In Artikel 2 (Abschnitt 6) werden auf der Grundlage sinnvoller Kriterien für die Auswahl von Szenarien für die Sicherheitsbewertung des automatisierten Fahrens relevante Testszenarien abgeleitet, in denen FRAM angewendet werden sollte. Artikel 3 (Abschnitt 7) verwendet FRAM in einem Überholszenario, um ein Modell zu erstellen, das die Muster der Unfallentwicklung und Unfallvermeidung versteht, um Empfehlungen für die Systemauslegung und wesentliche Erkenntnisse für den Validierungsprozess von automatisierten Fahrzeugen zu liefern. Artikel 4 (Abschnitt 8) bewertet das Modell und die Methode im Hinblick auf Validität, um die Glaubwürdigkeit der Ergebnisse bzw. die Anwendbarkeit zu beurteilen. In Abschnitt 9 werden die Forschungsergebnisse der Artikel 3 und 4 unter Verwendung eines rein funktionsbasierten Validierungsansatzes wieder aufgegriffen, um ein besseres Verständnis für die Glaubwürdigkeit des FRAM-Modells zu erlangen und erweiterte Muster zu identifizieren, die zur Systemverbesserung genutzt werden können. Schließlich werden in Abschnitt 10 die Ergebnisse in Bezug auf Systemdesign und -validierung, Methodenevaluierung und industrielle Anwendung diskutiert.

Die Forschungsergebnisse zeigen, dass Resilience Engineering, einschließlich FRAM und Safety-II, ein unschätzbarer und wesentlicher fehlender Ansatz für die Bewertung der Sicherheit des automatisierten Fahrens im Straßenverkehr ist, vor allem durch die Untersuchung von Interaktionen im Hinblick auf Emergentismen. FRAM ist ein vielversprechender Ansatz, der auf das Wesentliche abzielt - auf die Identifizierung der Muster, die die Anpassungsfähigkeit des Systems erleichtern, die für eine sichere und effiziente Leistung in soziotechnischen Systemen unverzichtbar ist. Ein solides Fundament in kleinem Maßstab, das sich mit dem richtigen Problem befasst, ist gelegt, wobei Verbesserungen und Erweiterungen in der Zukunft erforscht werden müssen, wie der Ansatz in der Industrie in großem Maßstab eingesetzt werden kann.

Keywords

Automated driving; Human driving, Safety and risk assessment; Resilience engineering; Functional resonance analysis method; Systems thinking and complexity; Safety-I vs. Safety-II; Socio-technical system, Overtaking manoeuvre; Validation

Contents

Nomenclature

Acronyms

ACC	Adaptive cruise control system
ADAS	Advanced driver assistance system
AI	Artificial intelligence
AUTOSAR	Automotive open system architecture
AV	Autonomous vehicle
CREAM	Cognitive reliability and error analysis method
CSTS	Cyber-socio-technical system
CWA	Cognitive work analysis
DARPA	Defense advanced research projects agency
DIKW	Data, information, knowledge, and wisdom
DL	Deep learning
DMS	Driver monitoring system
EAST	Event analysis of systemic teamwork
ESC	Electronic stability control
ETTO	Efficiency-thoroughness trade-off
EV	Ego vehicle
FiF	Function in focus
FMEA	Failure mode and effects analysis
FMI	Functional model interpreter
FMV	Functional model visualiser
FRAM	Functional resonance analysis method
FTA	Failure tree analysis
FVSRM	Functional variability-system resonance matrix
GNSS	Global navigation satellite systems
HABA-MABA	Humans-are-better-at/machines-are-better-at
HAD	Highly automated driving (SAE-Level 4/5)
HAV	Highly automated vehicles (SAE-Level 4/5)
HFE	Human factors and ergonomics
HMI	Human-machine interface
IMU	Inertial measurement units

ISO	International organisation for standardisation
JCS	Joint cognitive system
KPI	Key performance indicator
LiDAR	Light detection and ranging
LKA	Lane keeping assistant
LoA	Levels of automation
LoDA	Levels of driving automation
LV	Lead vehicle
ML	Machine learning
Net-HARMS	Networked hazard analysis and risk management system
ODD	Operational design domain
OuT	Object under test
OV	Oncoming vehicle
Radar	Radio detection and ranging
RE	Resilience engineering
REA	Resilience engineering association
RPG	Research-practice gap
RV	Rear vehicle
R&D	Research and development
SAE	Society of automotive engineers
SAFE	Situative Anforderungsanalyse von Fahraufgaben
SECI	Socialisation, externalisation, combination, and internalisation
STAMP	Systems-theoretic accident model and processes
STPA	Sytem-theoretic process analysis
SOTIF	Safety of the intended functionality
STS	Socio-technical system
ToA	Type of automation
TTC	Time-to-collision
UI	User interface
UX	User experience
VISSIM	Verkehr In Städten - Simulationsmodell
V2X	Vehicle-to-anything communication
WAD	Work-as-done

WAI	Work-as-imagined
WYFIWYF	What-you-find-is-what-you-fix
WYLFIWYF	What-you-look-for-is-what-you-find

Measures

DLFCV	Downlink functional coupling variability
FV	Functional variability
GSV	Global system variability
OFCV	Overall functional coupling variability
SR	System resonance
ULFCV	Uplink functional coupling variability
WaD	Weight as downstream
WaU	Weight as upstream

1 Introduction

“The question is no longer whether one or another function can be automated but, rather, whether it should be.”
– Wiener & Curry, 1980, p. 995 –

The vision from the automation of the driving task up to autonomous driving has always been a well-known human dream. Actually, it is now within reach: significant progress in the technical realisation of such vehicles is made, and numerous promises and positive marketing regarding an imminent market launch are usually reported. The first fully automated vehicles are predicted to be on roads by 2030 (Ertrac, 2017).

The automation of driving started in the 1950s. The General Motors Research Lab developed ideas on how driving on highways could initially be automated, whereby a combination of vehicle technology and infrastructure measures (e.g., the vehicle detects magnets inserted in the roadway) was considered promising due to the limiting computer performance at that time. (Fenton, 1970) In the 1970s and 1980s, Japanese groups researched the detection of lanes and objects with imaging cameras, which led to automated vehicle guidance (Tsugawa, 1994). According to today's standards, the results of that time correspond to an adaptive cruise control system (ACC) and a lane-keeping assistant (LKA) at very low speeds (Matthaei et al., 2015). In the USA, a step forward was crossing the USA by the test vehicle NavLab 5 in 1995. The assistance system took over the lateral guidance for 4.500 of the 4.587 kilometers driven on American highways (Pomerleau & Jochem, 1996). The PROMETHEUS project (PROgraMme for a European Traffic of Highest Efficiency and Unprecedented Safety, 1987-1994), funded by the European Union, developed similar systems that led to very efficient vehicles. This project culminated in a journey from Munich to Odense by the test vehicle travelling successfully 95% of the total of 1.758 kilometres at speeds up to 180 km/h through automation of both longitudinal and lateral guidance in 1995 (Matthaei et al., 2015). In addition, lane changes were initiated by the safety driver and then carried out automatically (Maurer, 2000). However, none of the previous vehicle prototypes corresponds to high or full automation (SAE J3016, 2021) due to a safety driver and a limited operational design domain (ODD) (Matthaei et al., 2015). Therefore, the USA's Defense Advanced Research Projects Agency (DARPA) aimed to develop driverless, autonomous vehicles (AVs) for military use at the beginning of the 2000s. To this end, the first DARPA Grand Challenge was held in 2004, in which driverless vehicles had to complete a course in the Nevada desert. Finally, the competition concept was expanded by DARPA 2007. Instead of driving through a desert, the driving missions were completed in a suburban-like environment with other road users to increase society's benefit. After initial difficulties, the competitions were considered a success because some teams were able to solve the set of tasks, giving the entire research community a boost that resulted in many vehicle automation projects (Matthaei et al., 2015): e.g., HAVE-it (2008-2011) (Hoeger et al., 2008), UR:BAN (2013-2016) (Bengler et al., 2018), AdaptIVe (2014-2017) (Langenberg et al., 2014), KoHAF (2015-2018) (ZENTEC Zentrum für Technologie, Existenzgründung und Cooperation GmbH), PEGASUS (2016– 2019) (German Aerospace Center [DLR]), interACT (2016-2020) (German Aerospace Center [DLR]), L3Pilot (2017–2021) (L3Pilot consortium), @City (2018-2022) ((At)City consortium), UNICARagil (2018-2023) (RWTH Aachen), and VVM (Verification Validation Methods) (2019-2023) (VVM consortium).

Today, according to the SAE J3016 (2021), we already have Level 2 functions in serial production, and in 2021 and 2022, the first vehicles with Level 3 functions for traffic jams on motorways by Honda and Mercedes entered the market (Slovik, 2021; Götze, 2021). It is an actual arms race between individual vehicle manufacturers competing with software companies. For instance, a robotaxi service called Waymo One in Phoenix and San Francisco offers fully autonomous rides in a limited urban area without any safety drivers physically on board but with monitoring operators who can intervene per teleoperation (Waymo LLC). Furthermore, in California, autonomous prototypes have been tested with safety drivers on various road types in recent years. Their incidents and experiences are reported annually in disengagement (a situation in which a driver takes over the control of the vehicle that can be initiated by the driver her/himself or by the system in case of system limits or malfunctions) and accident reports (cf. DMV California). They reported that between 2014 and 2017, twelve manufacturers tested 144 AVs, driving a cumulative 1.116.605 autonomous miles, and reported 5.328 disengagements and 42 accidents involving AVs on public roads (Banerjee et al., 2018). Banerjee et al. (2018) found that compared to drivers, AVs perform 15 - 22 times worse in accidents per mile, and 64% of disengagements resulted from problems in decisions made by the machine learning system. However, there is still much to optimise, as current accident rates and disengagement issues show (Dixit et al., 2016; Favaro et al., 2018, 2017). Especially, rather the entire AV system, which is still in a “burn-in” phase (Banerjee et al., 2018), and not the individual components must be safe.

The historical outline to the present day shows that the technical development of automated vehicles is well advanced, which is reflected in real hype. However, a valid concept for approval of such vehicles is still missing, which could slow down further development and initial euphoria. According to Winner (2016), the broad use of AVs in public road traffic will not be achieved until this problem is solved in an accepted form. Due to recent events of automated vehicles, such as the fatal accidents of a Tesla in 2016 (Boudette, 2017) or an Uber vehicle in 2018 (Griggs & Wakabayashi, 2018), the question of the safety of highly automated vehicle systems is more important than ever.

Experience from other industries, such as aviation, shows that automation may cause as many problems as it solves (Kyriakidis et al., 2019). Also, there is considerable knowledge about the promises and problems that can be extrapolated from the aviation domain to automated driving (Billings, 1993; Stanton & Marsden, 1996; Wiener & Curry, 1980). According to Walker et al. (2015), the following four potential issues for the automation of road vehicles can be anticipated based on hard-won lessons learned from automation in aviation: shortfalls in expected benefits, problems with equipment reliability, training and skills maintenance, and error inducing equipment designs. Based on this, Müller (2018) pointed out to the automotive industry that automation in aviation has been researched and used for decades and, in addition to positive effects, repeatedly leads to many problems. Therefore, Müller's advice to the automotive industry is the following: not to automate mindlessly and substitute the driver, but rather to support her/him, not to increase complexity unnecessarily, and to take more account of human and machine strengths and weaknesses and reasonably integrate them. However, is it possible to compare these two different transport systems and thus transfer the aviation automation challenges to road vehicle automation?

The air transport system and flying an aircraft fundamentally differ from the road transport system and vehicle driving. Despite these differences, it is assumed that the main findings concerning challenges of automation can be transferred between these transport systems due to similarities in terms of purpose (mobility and transport), physical and mental modes of operation (e.g., steering, monitoring), and relevant societal risks (Papadimitriou et al. 2020). First, the air traffic system is much more restricted than the road system due to separated and externally monitored airspaces. In addition, there are fewer concurrent traffic users in a much larger space, except in proximity to a terminal on the ground. There is, therefore, literally more collision-free space available in the airspace. Moreover, the diversity between air traffic participants is much smaller than in road traffic. Additionally, the tasks of a pilot are less time-critical due to a more significant time window for reaction time than when driving a car. Based on the aspects above, the air traffic system is less complex overall. Second, the quantity and quality of training required for operators also differ. Pilots are specially trained and highly skilled people for piloting an aircraft, whose handling of automation can be precisely controlled and trained. In contrast, driving a road vehicle is carried out by many different people with various degrees of training and skills, whereby targeted and comprehensive training for handling automation is impossible. (Papadimitriou et al. 2020; Ständer, 2010; Wachenfeld & Winner, 2016)

In this context, when considering that the transport system in which an aircraft operates is less complex than the road system and that automation is currently reaching its limits even in aviation, it becomes clear what a huge challenge we face regarding safe, highly automated vehicles (HAVs).

1.1 What is automated driving?

In general, automation refers to the full or partial replacement of a function by a machine previously carried out by the human operator (Parasuraman et al., 2000). Therefore, automation is not all or none but somewhat varies across a continuum of types and levels of automation according to the classification by Parasuraman et al. (2000) (see Figure 1). The type of automation (ToA) can be attributed to the simple four stages of human information processing: sensory processing, perception/working memory, response selection, and response execution. Parasuraman et al. (2000) translate these stages into four ToA or classes of functions that can be automated: information acquisition, information analysis, decision selection, and action implementation, respectively. The generic levels of automation (LoA) represent the extent to how much of a function or ToA is automated based on ten levels describing the allocation of the task from no allocation as manual control towards full allocation as complete autonomous control by automation (Sheridan, 1992). A specific system can involve automation of the ToA at different levels. For example, system A could be designed to have high LoA at all four ToA. In contrast, system B shows moderate acquisition automation, moderate to low analysis automation and decision automation, and low action automation (see Figure 1).

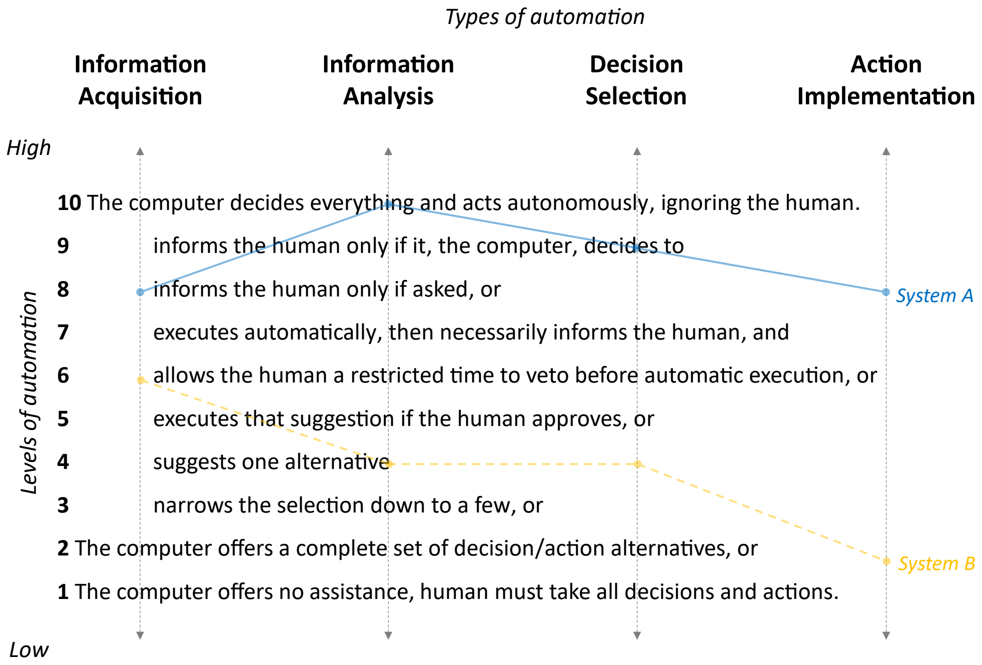

Types of automation
Information Acquisition
Information Analysis
Decision Selection
Action Implementation
High
Low
Levels of automation
10 The computer decides everything and acts autonomously, ignoring the human.
9 informs the human only if it, the computer, decides to
8 informs the human only if asked, or
7 executes automatically, then necessarily informs the human, and
6 allows the human a restricted time to veto before automatic execution, or
5 executes that suggestion if the human approves, or
4 suggests one alternative
3 narrows the selection down to a few, or
2 The computer offers a complete set of decision/action alternatives, or
1 The computer offers no assistance, human must take all decisions and actions.
System A
System B

Figure 1 The model of different types of automation by Parasuraman et al. (2000) combined with the levels of automation by Sheridan (1992). Examples of two systems with different automation levels across the four information processing stages are also shown.

In terms of automated driving, we have to examine the driving task itself closely. According to Geiser (1985), the primary, secondary, and tertiary driving tasks can be distinguished. The primary driving task requires the driver to keep the vehicle on the course at a specific speed. Here, three hierarchical levels of the driving task must be considered: navigation (strategic selection of driving route), guidance (tactical selection of maneuvers and trajectories, and stabilisation (operational control of the vehicle in the form of acceleration and steering). On a temporal scale, these levels take place at several seconds to hours, 2-15s, and 100-300ms, respectively (Bubb, 2015). Additionally, secondary driving tasks support the primary driving task and are necessary for dependency on the traffic and environment situation, e.g., the driver informs other road users about her/his intentions using the indicators or the lights, and windshield wipers are activated in reaction to lightning or weather conditions. Furthermore, the driver can optionally engage in tertiary driving tasks which do not relate to the actual driving task but rather enable an increase in comfort (e.g., seat position and air condition) or serve as information, communication, and entertainment purposes, such as taking a call or turning on the radio. According to Bubb (2015), the relationship between the different driving tasks and the dynamic interaction between the driver, vehicle, and environment can be described as a simplified closed feedback loop, as illustrated in Figure 2. The starting point is the driving task ending in a driving result. In terms of the terminology of control engineering, these are associated with a command signal and a tracking signal, respectively. The transfer process of the command to the tracking signal is accomplished by the controller (i.e., the driver) and the controlled system (i.e., the vehicle). The goal is to minimise the difference between the command and tracking signal. The difference is calculated at sum points, represented as circles, where command signals with a positive sign and tracking signals with a negative sign are fed in. Finally, the feedback of the result closes the loop characterising the control cycle. External influences, e.g., environmental conditions like road layout, other road users, or weather, can disturb the control process. Further disturbing variables could be driver-related factors, such as fatigue or workload, or vehicle-related factors, e.g., a burst tyre or an engine failure.

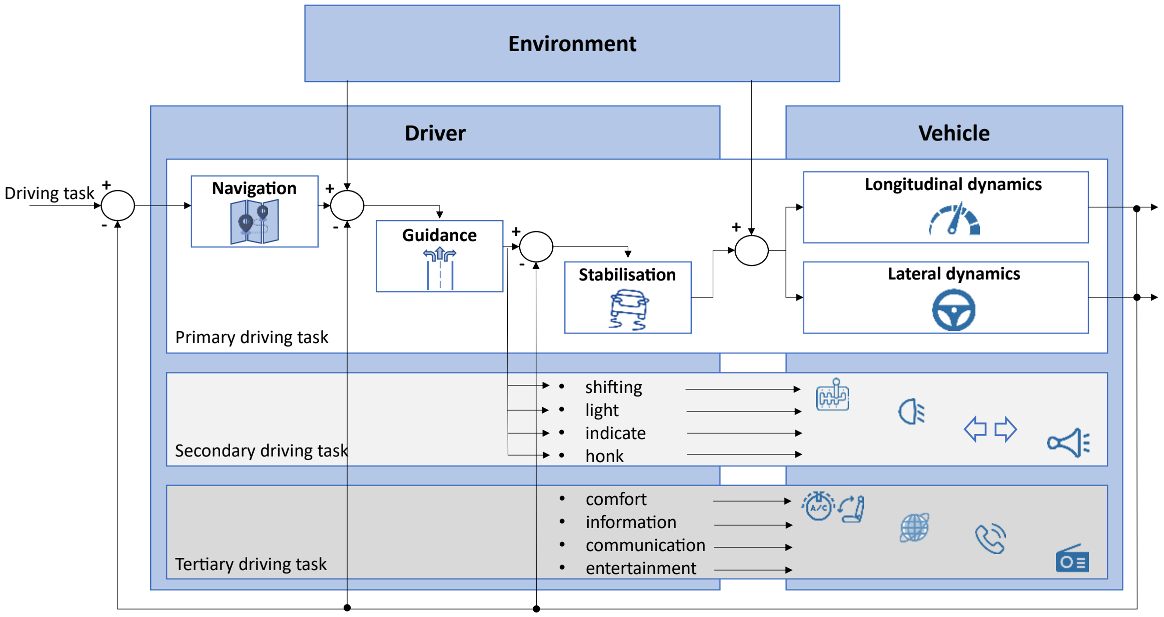
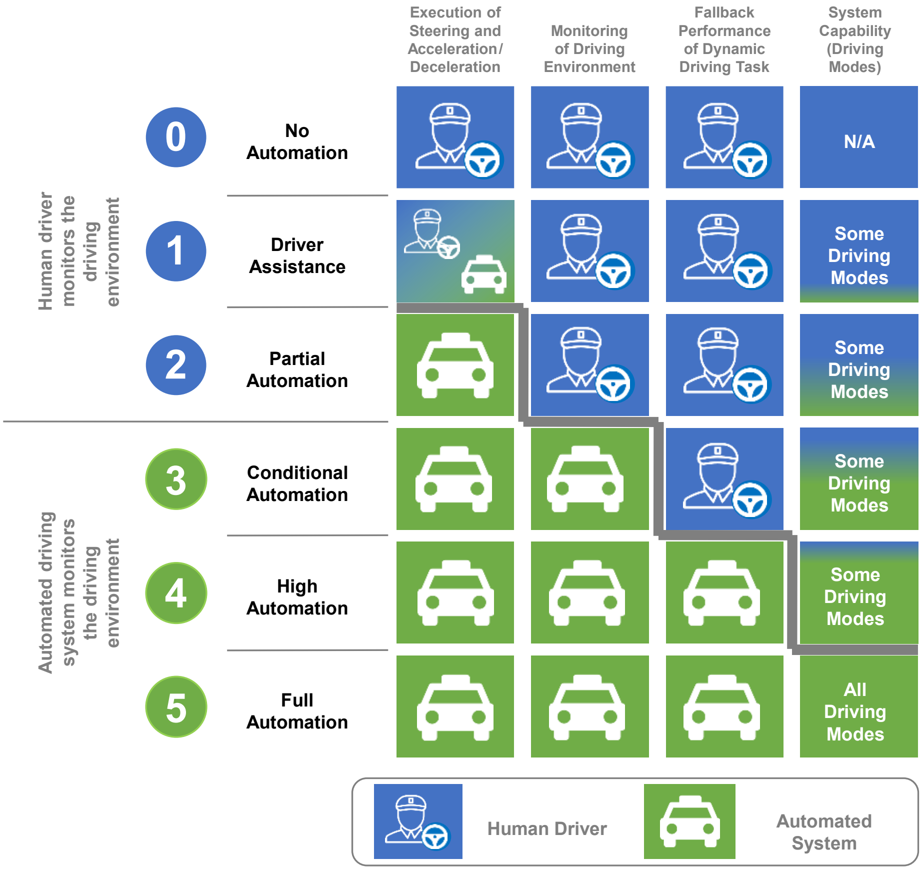


Figure 2 The driving tasks in the driver-vehicle environment feedback loop adapted from Bubb (2015).

In order to define and differentiate the responsibilities between the driver and the automation with regard to the primary driving task, the Society of Automotive Engineers (SAE) introduced six levels of driving automation (LoDA) (SAE J3016, 2021) which are commonly accepted and used in the industry. These LoDA are distinguished as depicted in Figure 3. At LoDA 0, the driver executes steering and acceleration of the vehicle, monitors the environment, and is the fallback solution for all driving modes. By further increasing the level of automation, the system performs more tasks of the driver and even in more driving modes. Systems up to LoDA 2 are already in serial production, and their safety has been successfully tested. These systems always support the driver in longitudinal and lateral control but do not replace her/him since the driver ultimately monitors the technical system and the traffic, serving as a fallback level. The decisive difference starts at LoDA 3, where the automation monitors the environment itself without drivers as supervisors, but as the last fallback level with a corresponding takeover request in case of system limits or malfunctions. The next major step follows in LoDA 4, where the machine itself acts as a fallback level, and the driver is completely removed from the driving task in limited driving modes. The only difference between LoDA 4 and 5 is the number of driving modes.



			Execution of Steering and Acceleration/ Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
Human driver monitors the driving environment	0	No Automation	🟦	🟦	🟦	N/A
	1	Driver Assistance	🟦 🟩	🟦	🟦	Some Driving Modes
	2	Partial Automation	🟩	🟦	🟦	Some Driving Modes
Automated driving system monitors the driving environment	3	Conditional Automation	🟩	🟩	🟦	Some Driving Modes
	4	High Automation	🟩	🟩	🟩	Some Driving Modes
	5	Full Automation	🟩	🟩	🟩	All Driving Modes

🟦 Human Driver 🟩 Automated System

Figure 3. Level of driving automation for road vehicles adapted from SAE J3016 (2021).

Besides, a special case exists concerning teleoperation where a remotely located human assists or operates an automated vehicle. The main idea is to overcome the potential limits of automated vehicles by providing a remote operator in case of system malfunctions or the system's incapability to cope with unknown or tricky situations when no driver is available inside the vehicle or no passenger is fallback-ready. The SAEJ3016 (2021) defined the terms remote assistance and remote driver. Remote assistance refers to the strategic and tactical level of the driving task. It provides information or advice (including pathways, revised goals, or classification of objects (Bogdoll et al., 2022) by a remotely located human to an automated vehicle operating in driverless mode to support trip continuation. Instead, in remote driving, a remote driver performs parts or all of the dynamic driving task in real-time, which belongs to the operational control, such as steering and acceleration.

Systems of LoDA 3 and above have not yet been successfully tested and do not exist as serial vehicles on public roads for a broad ODD, except narrowed ODDs such as parking, Waymo's robotaxi service in a limited urban area, or the limited functions by Honda and Mercedes, which are only used for traffic jams on motorways up to speeds of 60km/h (Götze, 2021; Slovik, 2021). In particular, LoDA 4 and 5 cause great difficulties regarding the proof of safety and are the focus of this thesis. These will therefore be referred to in the following as highly automated driving (HAD). The difficulties concerning the safety validation of HAD are explained in more detail in Section 2.

The introduction of driving automation cannot be considered in isolation from the driver due to mixed traffic, representing a long transition phase during which automated vehicles with varying degrees of automation and manually driven vehicles share the road (Bansal & Kockelman, 2017). In particular, the driver will still play a relevant role even in HAD (Christoffersen & Woods, 2002; Gasser et al., 2015). Alternatively, in other words, automated driving is not an "all or nothing" approach to human control, as implied by the LoDA (Steckhan et al., 2022). In fact, the LoDA are quite technology-driven and do not represent a driver-centric perspective (Noy et al., 2018). Rather user interventions have to be considered as well, which is also in line with the LoA and ToA. Thus, the concept of function allocation, which is closely related to the automation issue, must also be introduced. Function allocation refers to the design choice of assigning specific functions or tasks to humans or machines to accomplish a system objective (Inagaki, 2003). Traditional strategies of function allocation include a) comparison-driven assigning the function to the most capable agent (either human or machine), b) technology-driven allocating to machines every function that can be automated, and c) economy-driven finding an allocation ensuring the most economical efficiency (Inagaki, 2003). However, these traditional strategies are static, meaning that a given function allocation exists at all times and occasions and lacks a comprehensive human factors viewpoint. In order to tackle these criticisms, two types of human-automation collaboration and cooperation can be differentiated: shared control and traded control (Sheridan, 1992; see Figure 4).

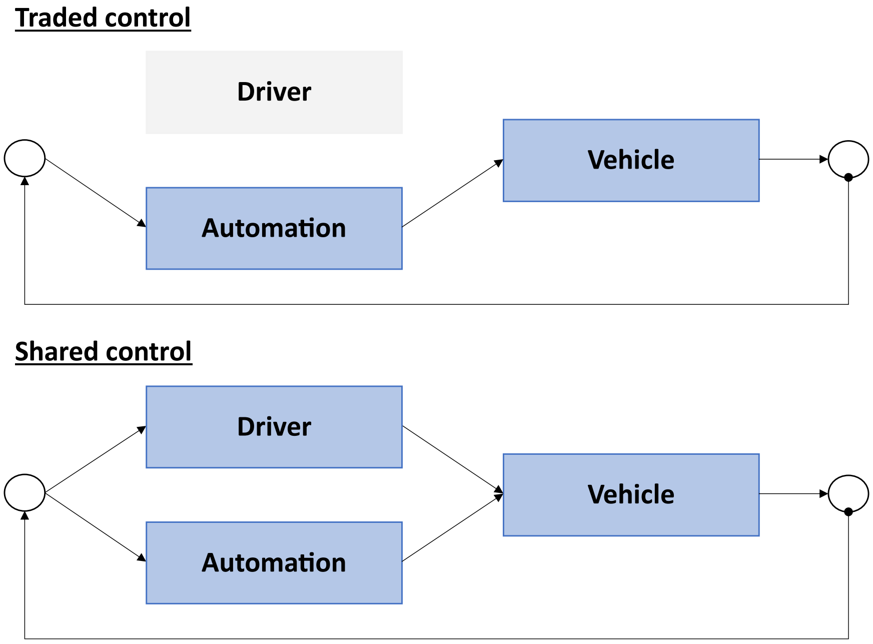


Figure 4. Conceptual control loops for traded and shared control adapted by Sheridan (1991).

Shared control means that the human and the automation work in a congruent collaboration simultaneously to achieve a single function (Sheridan, 1992). Two main types of shared control can be distinguished (Abbink et al., 2012): shared control involving a physically coupled interaction between an input device and vehicle or robot (called haptic shared control) and shared control involving a physically decoupled interaction (also called: blending, indirect, or input-mixing shared control). H-Mode is one example of a haptic shared control concept in driving inspired by the cooperation between rider and horse (Flemisch et al., 2014). The result of this approach is vehicle control on the stabilisation driving level in the form of a multi-modal combination of the driving automation system's intent and driver input via an active interface and fluid transitions between two levels of automation: tight rein (assisted) and loose rein (highly automated) (Flemisch et al., 2014; Kienle et al., 2009). In principle, according to the design and effect space of shared control and human-machine cooperation conceptualised by Flemisch et al. (2019), shared control is also possible on the guidance and navigation level, e.g., by maneuver-based driver-vehicle cooperation (Franz et al., 2012; Kauer et al., 2010; Walch et al., 2016; 2019). In contrast to shared control, traded control means that either the driver or the automation is responsible for a function, and their role can occasionally alter in time (Sheridan, 1992). The LoA combined with the ToA, and the LoDA, align with the concept of traded control but lack the concept of shared control. However, the LoDA only cover “who does what”, whereas the LoA and ToA also reveal “who does what and how”. One missing perspective is the adaptive function allocation, also called adaptive automation, enhancing a “when” (Inagaki, 2003). Adaptive function allocation is dynamic in nature but assumes explicit criteria to determine which functions should be reallocated, when, and how. The criteria involve various factors, such as environmental changes, task saturation of operators, and performance of operators (Inagaki, 2003).

1.2 Why automated driving?

Common reasons to automate a task or function formerly executed by a human can be found in the literature. These reasons can be classified into two main groups: typical reasons in practice and reasons related to a human factors perspective. According to Wickens et al. (2003), possible reasons from a human factors point could be:

- when a task is *dangerous*, e.g., teleoperated robots can be used when operating in hazardous environments or with hazardous materials like locating land mines
- when a task is *impossible* in the way that human capabilities are exceeded at both physiological and cognitive levels, e.g., disabled people/gravity conditions or complex mathematic calculations
- when a task is *tedious* and thus a *burden* for the human operator, e.g., repetitive tasks like assembly line works or vigilant monitoring
- when a task is *error-prone*, e.g., performing repetitive tasks under time pressure or stressors
- to *aid* by combining human and machine strengths, e.g., using the HABA-MABA (humans-are-better-at/machines-are-better-at) list by Fitts (1951) or its updated version by De Winter & Dodou (2014)

Instead, typical reasons in practice are associated with technology-driven and economic reasons (Wickens et al., 2003), among others:

- simply a solution to automate if *technically feasible*
- *efficiency and cost reduction* by reducing human resources and increasing the production rate

- *a reliable and repetitive output* by eliminating human variability

Unfortunately, the typical reasons in practice are usually preferred over the human factors-related reasons in an inappropriate proportion, which probably results in safety issues.

In particular, the introduction of automated driving to road traffic is motivated by several expected, beneficial outcomes (Maurer et al., 2015; Watzenig & Horn, 2017b): First, traffic safety was improved by three major safety strategies, including engineering, enforcement, education (Hughes et al., 2016), and their intertwinings. However, over 1.2 million people die each year on the world's roads, and between 20 and 50 million suffer non-fatal injuries (World Health Organisation, 2020). These are still high numbers that need to be improved. The human in his role as a driver is frequently assumed as the leading cause of accidents, claiming that human error causes approximately 90% of road crashes, e.g., (Dingus et al., 2016; Hendricks et al., 2001; Otte et al., 2009; Singh, 2015). Even though driver assistance systems have already mitigated human error (Golias et al., 2002), it is assumed that the automation of the driver's tasks will alleviate it further still because primary causes of accidents, such as speeding, misjudgment of distances, or distraction (Dingus et al., 2016; Klauer et al., 2006; Reichart, 2000), may be eliminated. Second, the act of driving can induce stress (Matthews et al., 1996) and has been linked to adverse effects on the driver's well-being and mood, particularly during routine drives like commuting (Roberts et al., 2011). Instances of frustration and aggression, commonly known as "road rage," can arise as a result (Dollard et al., 1939; Joint, 1995; Parker et al., 1998; Shinar, 1998). Automating this activity can potentially improve individuals' overall comfort and well-being. Additionally, it would allow drivers to utilise their travel time more efficiently by engaging in non-driving-related tasks, such as reading, watching movies, sleeping, or using electronic devices like laptops, tablets, and phones (Feldhütter et al., 2016; Gold et al., 2016; Hecht et al., 2019). Apart from enhancing comfort, automated driving also aims to enhance the mobility of individuals with physical impairments and age-related mobility limitations, granting them independence and fostering inclusivity in their daily lives (Shergold et al., 2016). Third, as the capacity of the traffic system can only be expanded to a certain extent, increasing traffic efficiency and optimising capacity utilisation become crucial goals. Thus, automated and connected vehicles could help to achieve this goal by reducing speed variability, route planning according to the current traffic, and more efficient driving (Friedrich, 2015; Tientrakool & Maxemchuk, 2011; Wagner, 2015). Furthermore, fewer traffic jams through increased traffic flow could reduce fuel consumption (Khondaker & Kattan, 2015; Wu et al., 2011) and air pollution (Bose & Ioannou, 2001).

Bengler et al. (2017) state that interdependencies between safety, efficiency, and comfort occur. For example, the increase in safety can also be based on the technical system's compliance with maximum speeds and speed-dependent minimum distances, which, however, can lead to a reduction in traffic flow in return (Shladover et al., 2012; Van Arem et al., 2006). However, Popiv et al. (2010) show that the driver can balance safety, efficiency, and comfort. Therefore, the positive expectations only represent potentials and hoped-for assumptions that must be explicitly demonstrated.

1.3 Basic technology of automated driving

The following briefly overviews the underlying technology and technical architecture of automated driving. A comprehensive description of the technical components of automated vehicles is beyond the scope of this thesis. Hence, a more detailed overview can be found, for example, in Watzenig & Horn (2017a), Winner et al. (2016), or Yeong et al. (2021). According to Velasco-Hernandez et al. (2020), in order to realise automated driving in road traffic, two main viewpoints can be considered: a technical perspective (see Figure 5), which incorporates the hardware and software components of the automated vehicle, and a functional perspective (see Figure 6), that describes the information processing stages as functional blocks and the flow of information from data collection to the control of the vehicle.

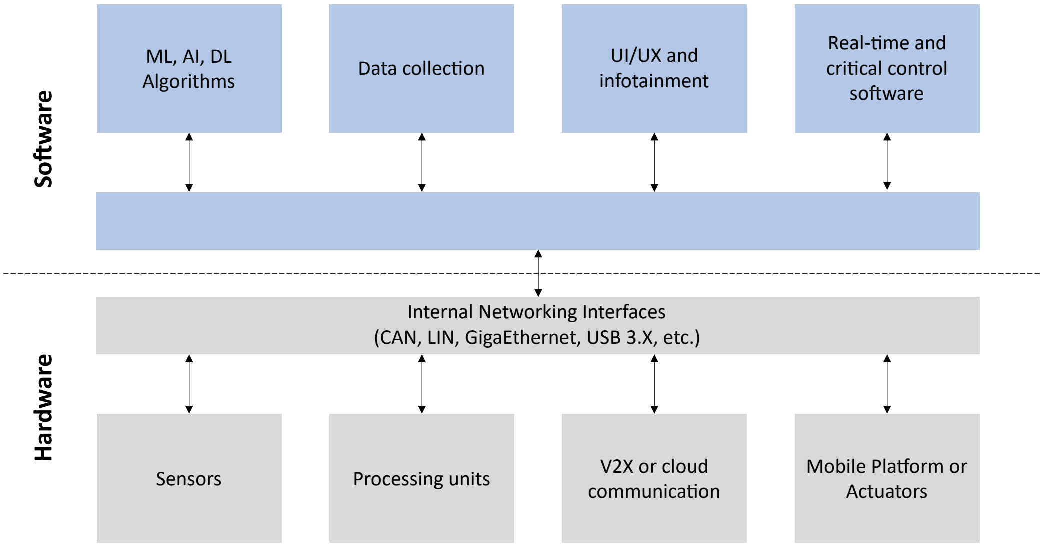

Software
ML, AI, DL Algorithms
Data collection
UI/UX and infotainment
Real-time and critical control software
Hardware
Internal Networking Interfaces (CAN, LIN, GigaEthernet, USB 3.X, etc.)
Sensors
Processing units
V2X or cloud communication
Mobile Platform or Actuators

Figure 5 Technical architecture for an automated driving system adapted by Velasco-Hernandez et al. (2020).

In the following, the technical view is described. The automated vehicle is equipped with several *sensors* used for internal and external monitoring to generate data as information. This information coming from the sensors is further gathered and processed by *processing units*. In addition to the data created by the vehicle, external data from the internet, other road users, or infrastructure are also available, known as *vehicle-to-anything communication (V2X),* expanding the coverage of vehicle sensors to exchange enhanced information. The hardware part also consists of the vehicle as a *mobile platform and actuators*. Finally, each hardware subsystem can exchange information through the *internal networking interfaces*. The software components include frameworks, libraries, and modules that support *Machine Learning (ML), Artificial Intelligence (AI), and Deep Learning (DL) algorithms* required for processing and understanding the data, drivers for *data collection* from the sensors, *user interfacing* (UI) and *user experience* (UX) through the *infotainment system*, and *real-time and critical software* for controlling actuators and monitoring the status of the vehicle. The software parts are inherently complex, requiring software frameworks and standards to enable such systems' successful development, management, and deployment. One example of software guidelines and frameworks is the AUTomotive Open System ARchitecture, AUTOSAR (see Staron & Durisic, 2017). Velasco-Hernandez et al. (2020)

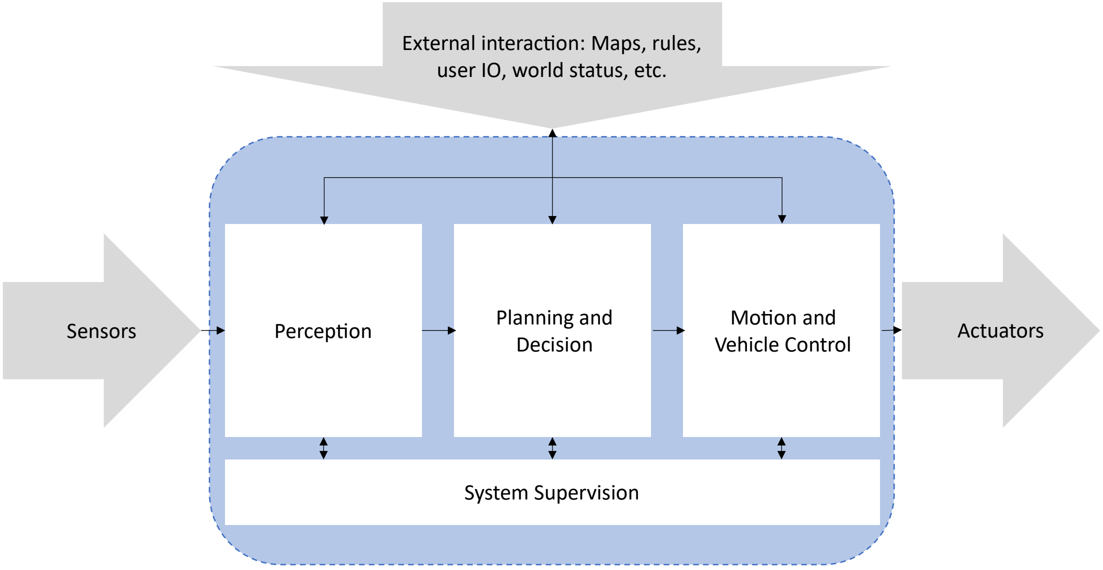

External interaction: Maps, rules, user IO, world status, etc.
Sensors
Perception
Planning and Decision
Motion and Vehicle Control
Actuators
System Supervision

Figure 6 Functional architecture for an automated driving system adapted by Velasco-Hernandez et al. (2020).

Regarding the functional view, four main functional blocks can be distinguished: *perception, planning and decision, motion vehicle control, and system supervision*. In the *perception* stage, data is received and fused from *sensors* or other sources, e.g., maps, to represent the vehicle status and the environment, mainly used for localisation, mapping, and object detection. Typically, sensors are categorised into proprioceptive and exteroceptive sensors. Proprioceptive sensors (e.g., Global Navigation Satellite Systems (GNSS) or Inertial Measurement Units (IMUs) are used as vehicle state sensing to get position, movement, and odometry information of the platform. Additionally, in-vehicle sensors can be used to determine the driver's or passenger's status and intention. Instead, exteroceptive sensors (e.g., Radar (Radio detection and ranging), LiDARs (LIght Detection And Ranging), cameras, and ultrasonic) monitor the environment to obtain data, e.g., the terrain, the road layout, and external objects. Velasco-Hernandez et al. (2020)

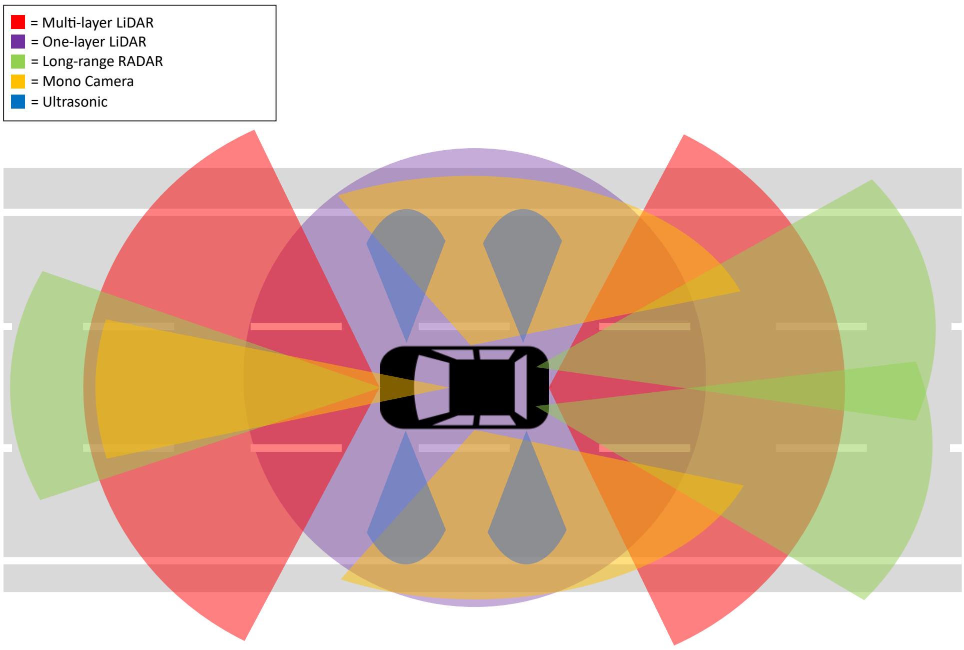


Figure 7 An example sensor configuration of an automated vehicle adapted from Aeberhard et al. (2015) and Wendt & Cook (2018).

Figure 7 shows an exemplary exteroceptive sensor configuration of four different sensor types following the redundancy principle. This means that most areas are covered by multiple sensor types in order to ensure robust and reliable sensory data because no sensor type works well for all tasks in all conditions. In the following, the basic functioning and application of the different and most common sensor types are described:

- Radar: Radar sensors operate on the principle of radiating electromagnetic waves, which are received as reflections by objects to establish a range information about target objects (Skolnik, 1962). In particular, the Doppler effect is used to determine the relative speed and relative position of detected obstacles (Shahian Jahromi et al., 2019). Long-range systems (77GHz) and short-range systems (24GHz) can be differentiated. Overall, radar is independent of light and weather conditions. In particular, long-range radars can detect objects up to 250m with a small spread in very adverse conditions where no other sensor works (Marti et al., 2019). Instead, short-range radars have a lower range but a higher spread. Potential difficulties of radar sensors are the sensible target reflectivity due to the heterogeneous reflectivity of different materials (Marti et al., 2019).

- LiDAR: These sensors are based on the principle of emitting pulses of infrared beams or laser light, which reflect off target objects (Li & Ibanez-Guzman, 2020). This is used for estimating object distances at relatively low distances compared to Radar but also generating a 3D representation of the surroundings as a point cloud providing a 360-degree detailed and accurate image (Campbell et al., 2018). However, LiDAR sensors are challenged by small and specular objects and are affected by adverse weather conditions like rain and fog (Marti et al., 2019).
- Camera: A camera works on the principle of detecting lights emitted from the surroundings on a photosensitive surface through a camera lens to create high-resolution images of the surroundings (Campbell et al., 2018). This enables the system, in addition to spatial and kinematic information, to identify semantic and qualitative information, e.g., road signs, traffic lights, road lane markings, gestures by humans, and shapes of other road users to distinguish, e.g., pedestrians, cyclists, motorcyclists, passenger car and trucks. Potential drawbacks are varying light and visibility conditions (Marti et al., 2019). In principle, monocular or binocular (stereo) cameras can be distinguished. When using stereo cameras, depth of field can be included. Infrared systems can be used for night vision (Punke et al., 2016).
- Ultrasonic: These sensors use ultrasonic waves to calculate the distance between the vehicle and the object, typically placed around the vehicle, for a redundant detection of very close objects ranging from 50 to 400cm (Paulweber, 2017). A typical application is assisted or autonomous valet parking.

A more detailed review of the advantages, drawbacks, and challenges of sensor technologies in terms of exteroceptive perception can be found in Marti et al. (2019).

Based on the *perception* stage, the *planning and decision* stage derives a driving strategy as the best possible trajectory that moves the vehicle safely, efficiently, and comfortably based on the current traffic situation and the goals or travel mission, including long-term plans (navigation from place A to place B on a global journey map) and short-term plans (guidance from waypoint to waypoint on a dynamic local map). This stage also includes information on *external sources* like traffic rules, map updates, user interventions, or speed limits. Finally, the *motion and vehicle control stage* is about sending the trajectory as movement commands to the platform and actuators for execution, considering their characteristics and limitations. These commands are on a stabilisation level, e.g., longitudinal speed, steering, and braking. Another functional block is a *supervision system* that monitors all automation system components, i.e., the hardware and software. The task is to ensure as a safety-critical system that possible malfunctions of hardware or software components do not harm humans, vehicles, or the environment. This is, for example, addressed by ISO 26262:2018 as a standard for functional safety in road vehicles. Velasco-Hernandez et al. (2020)

1.4 Idiosyncrasies of the safety argumentation of automated driving

Three main safety strategies or tools, including engineering, enforcement, and education, can be applied to the traffic system components to enhance road safety. Engineering includes actions to prevent an accident (primary safety, e.g., ACC) or mitigate the severity of an accident (secondary safety, e.g., airbags). A distinction should be made between measures concerning the vehicle or the infrastructure (e.g., road layout or crash barriers). Enforcement is about informing traffic participants of compliance with traffic rules and punishing their violations. For example, radar controls are used for speed measurements, resulting in potential financial penalties or disqualification from driving. Education describes interventions (e.g., safety training or targeted public campaigns) to improve the knowledge of necessary rules and their respective context, assessing the traffic situation, and skills for passive and active traffic participation. (Bubb, 2015).

These safety strategies were applied in the past to improve road safety. Consequently, traffic fatalities dropped from over 21,000 in 1970 to almost 3,000 a year today despite the huge increase in the number of motor vehicles registered in Germany (see Figure 8). Figure 8 illustrates the safety measures by the horizontal grey, blue, and green boxes, for example, the mandatory seat belt use in 1976, the front airbag in 1980, and electronic stability control (ESC) in 1995. It should be noted that the statistics for Germany only exemplify the subsequent explanations and arguments. Hence, it is also relevant for the settings of other countries. Unfortunately, the number of fatalities has reached a low plateau since 2012. The current approach and the stated countermeasure to further decrease the fatalities lie in automating the driving task with the long-term goal of fully automated driving, thus eliminating the driver. This approach follows the safety strategy of engineering, specifically with the vehicle's primary safety. It can also be attributed to infrastructure if V2X communication is considered, which, however, is useless in mixed traffic, including, for example, vulnerable road users and vehicles with different automation levels.

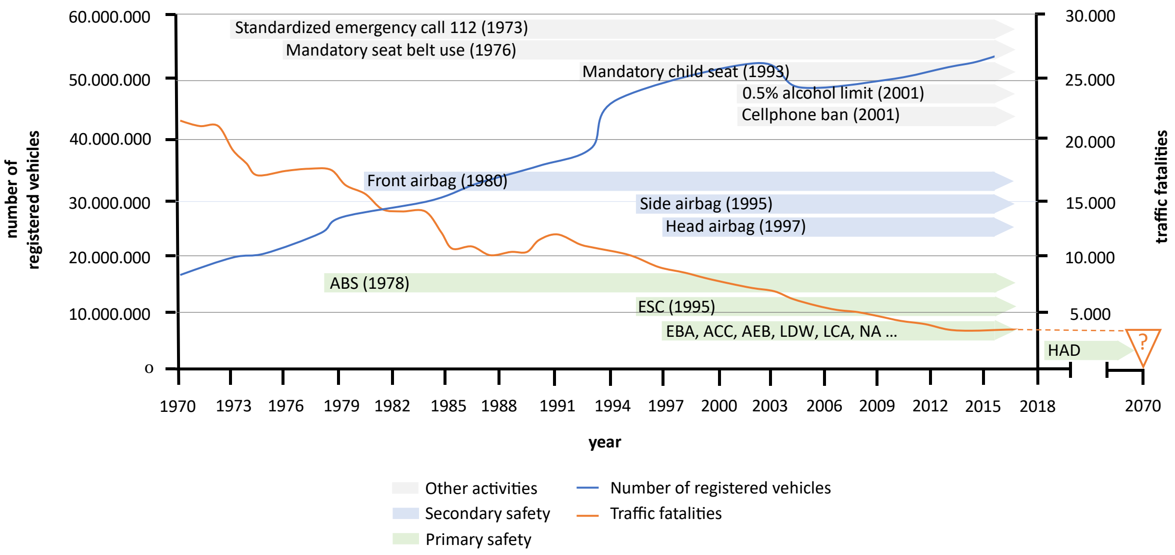


Figure 8. Drop in traffic fatalities (orange line) due to safety improvements (horizontal boxes) despite the increase in registered motor vehicles (blue line) in Germany. The safety improvements belong to other activities (grey), secondary safety (blue), and primary safety (green), according to Grabbe et al. (2020b).

The argument for increased automation in the driving task is often accompanied by the reasoning that the human in her/his role as a driver and the main cause of accidents could be removed from the system. Consequently, the number of accidents would fall sharply, assuming that human error is entirely an inappropriate behavior on the driver's part and that technology is error-free (Noy et al., 2018). However, this thinking pattern represents a persistent oversimplification fallacy called substitution myth (Woods & Dekker, 2000; cf. Drösler, 1965). Here, the concept of cause falsely links the logic of a clear causal link (Bengler et al., 2017; Grabbe et al., 2020b):

The *driver* causes an *accident*.

If the cause *driver* is removed, then the effect *accident* disappears. However, this cause-effect relationship only applies to mono-causal and resultant events (see Figure 9). A road accident is a rare, poisson-distributed, multi-causal, and emergent event. Therefore, it is crucial to keep in mind that the driver is involved in a road traffic accident as one of several interacting factors (e.g., other road user, road layout, environment conditions, or vehicle components) and has not prevented this accident at the last moment (see Figure 9):

The *driver* is involved in an *accident* in addition to *other interdepending factors*.

If the *driver* is removed, the *other factors* still apply to the remaining participants and elements in the system. Removing the *driver* would eliminate the negative and positive contributions that the driver brings into traffic (see Figure 9). This results in a differentiated consideration of the mechanisms of accident development. Above all, this also addresses the mechanisms of accident prevention. In addition, the fact that not only is the *driver* removed from the system but is replaced by the introduction of an *automated vehicle* with currently unknown system consequences has to be considered. All in all, the driver is both an active or passive participant in an accident and an accident avoidance and compensation element in the same system (Bengler et al., 2017).

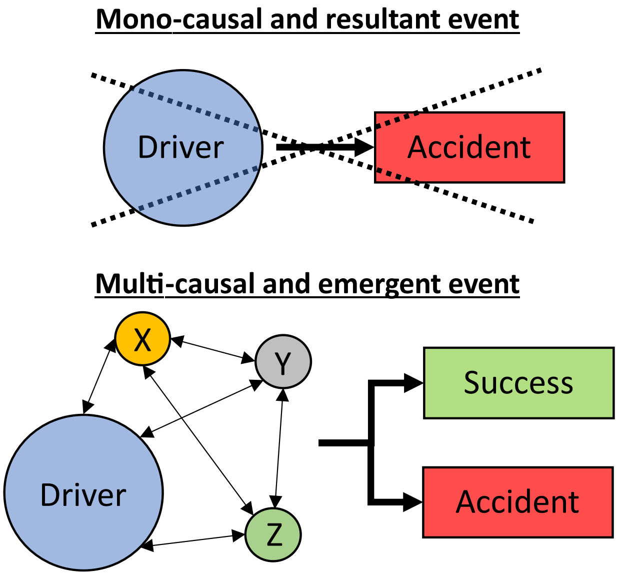

Mono-causal and resultant event
Driver
Accident
Multi-causal and emergent event
X
Y
Z
Driver
Success
Accident

Figure 9. Simplified scheme of cause-effect relationships distinguishing between mono-causal and multi-causal events applied to the road traffic system.

Ultimately, the automation system must have acceptable performance in situations where the driver is error-prone, but especially in situations where the driver usually performs well (Bengler et al., 2017). Therefore, recent accidents must be prevented, and no new accidents must be caused. There is no doubt that automation systems, which can compute fastly, consistently, and precisely, can eliminate accidents related to typical driver errors, such as speeding, misjudgment of distances, or distraction (Dingus et al., 2016; Klauer et al., 2006; Reichart, 2000). However, the current characteristic of drivers to adapt to changing system conditions to compensate for adverse road behaviors and conditions and to prevent accidents constitutes the more challenging task for automation. For example, the work of Reichart (2000) shows that human errors occur with very low probabilities of 10^{-3} to 10^{-4} in obscuring objects, interpretation, or steering errors. Moreover, based on the facts that drivers have a fatal accident every 90 million km, make 125 observations, and make 12 decisions on average per driven kilometre, Fastenmeier (2015) reasoned that a wrong driver decision leading to a fatal accident would be taken after about 10 billion observations and 1 billion decisions. Shladover & Nowakowski (2019) made similar calculations for road traffic in the USA. These numbers give an idea of the human's high performance in the driving task subtasks, demonstrating the huge challenge for automation to achieve or even exceed the human driving performance.

In fact, it is not clear that HAD will ever be safer than an experienced middle-aged driver representing the baseline (Sivak & Schoettle, 2015) as neither eliminating driver errors necessarily eliminate other factors contributing to accidents (Noy et al., 2018) nor automation systems are undoubtedly safe and reliable (Martens & van den Beukel, 2013; Schoettle & Sivak, 2015). Thus, it is claimed that fully automated vehicles may never operate at acceptable levels. Hence, automation should be used on specific routes, under specific conditions, and for target applications (Kyriakidis et al., 2019). Nevertheless, it is certain that the attribution of driver, vehicle, and environmental causes will significantly change (Noy et al., 2018). Thus, probably not the frequency of accidents but rather the quality of accidents and their black spots will alter because human accidents will be partly prevented (Gasser et al., 2012). However, new automation risks arise, and the distribution of different severity classes is unknown (Wachenfeld & Winner, 2016). Additionally, accidents are assumed to be emergent features of the road system, which is why accidents will inherently occur even without driver involvement (Bengler et al., 2017). Further, it is almost neglected that every scenario offers different potentials for automation as accident black spots (Maier, 2013; Gründl, 2005) can show, and drivers benefit unequally from automation and deal with potential side effects of automation in various ways as accident-prone drivers indicate (Das et al., 2015; Visser et al., 2007).

As we have previously seen, the expected safety benefits of automated vehicles are highly questionable. When not considering the safety aspect of automated vehicles from a more differentiated view, i.e., a more human-centric as well as systemic perspective, adverse automation surprises, as happened in other domains such as aviation, e.g., (Billings, 1993; Stanton & Marsden, 1996; Wiener & Curry, 1980), will probably occur due to safety blind spots (Noy et al., 2018) and could ultimately develop into a “showstopper”. The safety blind spots mainly constitute ironies and pitfalls of automation, acceptance issues from different stakeholders and ethical considerations, and a systemic consideration of automated vehicles in the context of socio-technical systems (STSs) or even in a cyber-physical world. These safety blind spots are explained in the following.

Ironies and pitfalls of automation can be seen as unintended system consequences of introducing automation. This manifests in the phenomenon that the demands on the driver are exaggerated rather than decreasing or resolving driver workload and vigilance. This is due to the paradox that the more advanced an automation system is (except full automation), the more crucial may be the role of the driver because routine tasks that the driver accomplishes typically well are automated, but the complex and challenging tasks are still left over to the driver. Several papers addressed these ironies and pitfalls (e.g., Bainbridge, 1983; Hancock, 2019; Fitts, 1951). As the driver is still needed, e.g., for supervision and as a fallback, technological changes lead to dynamics and adaptations by the driver collaborating with automation. This may result in negative side-effects, such as risk adaptation in the form of risk or task difficulty homeostasis (Wilde, 1982; Fuller, 2005), automation surprises (e.g., Sarter et al., 1997), or the out-of-the-loop performance problem (e.g., Endsley & Kiris, 1995) resulting in deskilling due to a lack of practice (e.g., Wiener & Curry, 1980), loss of situational awareness and mode confusion (e.g., Sarter & Woods, 1995; Wickens, 1995), complacency or overtrust and mistrust leading to misuse/abuse and disuse (e.g., Lee & See, 2004; Parasuraman & Riley, 1997), and vigilance problems or boredom leading to frequent non-driving-related activity engagement which in turn causes distraction (e.g., Carsten et al., 2012; Saffarian et al., 2012).

In addition, the acceptance of automated vehicles in public transport by different stakeholders is a critical element in the form of a risk-benefit relationship (Starr, 1969) to define an acceptable level of risk or safety (e.g., Liu et al., 2019; Shariff et al., 2021). This results in a social risk constellation as different groups like decision-makers, regulators, politicians, drivers/passengers, and other road users are affected and profit differently from automated driving. Overall, an active and passive confrontation with risks have to be distinguished (Grunwald, 2016): active users as drivers/passengers who can decide whether or not to take the risk arising from automated vehicles, and passive road users are exposed to risks that they cannot avoid or only avoid with considerable effort and drawbacks. Active users will have a higher risk level than passive road users as they have higher exposure as long as the number of automated vehicles is low. However, they probably benefit most from a functionality perspective (Wachenfeld, 2017). In addition to benefits and risks, acceptance also depends on people's values, which are influenced by, e.g., age, gender, or culture. Furthermore, a particular concern in terms of acceptance is ethical trade-offs. An issue is an algorithm deciding between affecting the passenger or others when facing an inevitable crash. This is generally defined as the *trolley problem* (Foot, 1967) or, more specifically, labeled as a *social dilemma* of AVs (Bonnefon et al., 2016). For example, in 2017 and 2020, the German Ethics Commission for Automated and Connected Driving and the EU Commission Expert Group, respectively, released 20 ethical guidelines (Bonnefon et al., 2020; Di Fabio et al., 2017) which serve as a guidance for basic ethical behavior of AVs. In order to make this debate more specific, Bonnefon et al. (2019) transformed the trolley problem into a statistical thought experiment adopting the theory *ethics of risk.* Building on this, Geisslinger et al. (2021; 2023) presented an ethical trajectory planning of AVs based on a risk-cost function. The aforementioned social risk constellation must be considered in this calculation process in the future.

Ultimately, current safety analyses primarily focus on the automated vehicle itself. However, Grabbe et al. (2020b) argue to broaden the focus more on a systemic level and view automated vehicles in an integrated STS or even within a cyber-physical world (Noy et al., 2018). In this context, the automated vehicle is just one element interacting with various elements in the entire system, influencing overall performance. For example, research by Preuk et al. (2016) and Ma & Zhang (2022) implies that in mixed traffic, manual drivers adapt their driving behaviour when encountering automated vehicles because those vehicles do not behave like regular drivers. For instance, drivers show more aggressive behaviour as bullying HAVs (Liu et al., 2020), exploit the HAV's defensive programming in time-critical situations (Trende et al., 2019), or accept shorter headway gaps for increased HAV penetration levels (Chityala et al., 2020). Safety is thus an emergent and complex system property resulting from the interactions among system components (Leveson, 2011) rather than the individual performance of system components following the traditional reductionist approach. These two fundamental concepts concerning safety are explained in more detail in Sections 2.5 and 3.

This systemic view can help identify potential conflicts in the flow of interactions and analyse trade-offs to reveal unforeseen adverse consequences to optimise the overall system design. According to Noy et al. (2018), the safety of automated vehicles is not just about the driver, designing the human-machine interface, and the driving task itself, it is more about all people in the system, creating value and the right level of trust and acceptance, and mobility within a cyber-physical world, respectively.

1.5 Thesis Outline

Sections 5 - 8 are based on a publication. They are attached in their original format in the Appendix B-E of this thesis.

As previously indicated, the safety of automated vehicles is controversial, and no proof of HAD has yet been provided. Therefore, Section 2 describes the challenges to assessing the safety of HAD. It begins with presenting the current test concept in automotive and shows HAD's differences and unique features. Further, the resulting approval trap (Winner, 2016) and associated common approaches to solving this problem are explained. Finally, a naive fallacy in these common approaches is pointed out, and a new perspective on safety, almost neglected, is presented.

Section 3 illustrates the historical development of the scientific study of safety and risk following different “ages”. This overview synthesises the evolution from traditional reductionist reasoning towards a complexity-oriented systemic approach based on resilience engineering.

Section 4 presents the research goals and questions that result from the new perspective regarding the safety assessment of automated vehicles. In addition, a structural overview of the research process is given, and how the objectives and research questions are answered in the publications and beyond.

Section 5 introduces the need for a systems approach, with particular attention to Safety-II, to address the safety assessment of automated driving. Especially the application of the functional resonance analysis method (FRAM, Hollnagel, 2012a) is discussed in a case study.

Section 6 addresses the scenario-based approach to overcome the approval trap by focusing on road safety mechanisms to facilitate the reduction of relevant scenarios for testing. Relevant scenarios are deduced based on reasonable criteria for scenario selection, classified in an abstract form, and presented using a systemic analysis method.

Section 7 analyses the contribution of automated driving to road traffic safety compared to a driver in an overtaking manoeuvre on a rural road using FRAM. Therefore, an in-depth instantiated FRAM model was developed and enhanced by a semi-quantitative approach combined with a Space-Time/Agency framework. Finally, general and specific system design recommendations and essential insights for the validation process are given.

Section 8 provides a formal approach to achieve and demonstrate the reliability and validity of an instantiated FRAM model. In particular, the predictive validity of the former developed FRAM model and its applicability are evaluated to assess the performance and value of the FRAM method with regard to the safety assessment of automated driving.

Section 9 revisits the research findings of Sections 7 and 8 using a pure function-based validation approach differentiating instantiations to get an enhanced comprehension with regard to the FRAM model's credibility and to identify enhanced patterns that can be used for system improvement.

Finally, the main results and conclusions, including limitations, recommendations, and future research, are discussed in Section 10**.** In particular, the individual research results are discussed on an abstract level across three dimensions: system design and validation, method evaluation, and industrial application. The main goal is to elucidate the potential benefit of FRAM and its reasoning for enhancing road safety in the context of automated vehicle introduction.

2 Safety Assessment Challenges of Automated Driving

> “We cannot solve problems by using the same kind of thinking we used when we created them."
>
> – Albert Einstein –

In general, the transition of a technical system from the development phase to serial production requires the release of this system (Felkai & Beiderwieden, 2011). The release only takes place when this technical system fulfils the previously defined requirements. In particular, the safety of people in road traffic must be met, the increase of which is one of the major drivers of vehicle automation. This results in the minimum requirement that the proposed long-term substitution of drivers in road traffic by automation does not reduce road safety. This objective should apply to the occupants and the entire transport system in which the automated vehicle operates. However, these requirements pose a great challenge with regard to the proof of safety for HAD. Why this is the case is discussed in the following subsections.

First, the current test concept and its premises for the argumentation of a safety validation are presented. Then, the differences and peculiarities caused by HAD are shown, which makes the current test concept unusable. Consequently, this results in the so-called approval trap, which is explained in more detail. Afterwards, common approaches to solve this problem are presented. Finally, a fallacy in all these common approaches is pointed out, and a new perspective on safety, almost neglected, is presented.

2.1 Current test concepts in the automotive industry

The following contents in Subsections 2.1 to 2.3 are essentially based on the statements of Wachenfeld & Winner (2016).

Current systems in serial production can be assigned to LoDA 0 to 2. In all these systems, the release concept is based on controllability by the driver: either to enable the driver to control the system or to restore controllability for her/him. According to ISO 26262:2018, controllability refers to the entire automation system-driver-environment interaction comprising:

- normal system use within system limits,
- usage at and beyond exceeding system limits,
- and usage during and after a system failure.

The driver as the fallback level is thus the basis for approving current vehicles. It must also be shown that the vehicle components do not exceed a specified maximum failure rate. The development and proof of the controllability for the driver are carried out according to the V-model (see Figure 10), whereby a distinction is made between the left descending branch of product development and the right ascending branch of verification and validation as a means of quality assurance. (cf. Wachenfeld & Winner, 2016) For quality assurance, a test concept is followed, which, according to Schuldt et al. (2013), includes the analysis of the test object, the test case generation, the test execution, and the test evaluation. The analysis and test case generation are carried out in the product development phase, whereas test execution and evaluation take place in the validation phase.

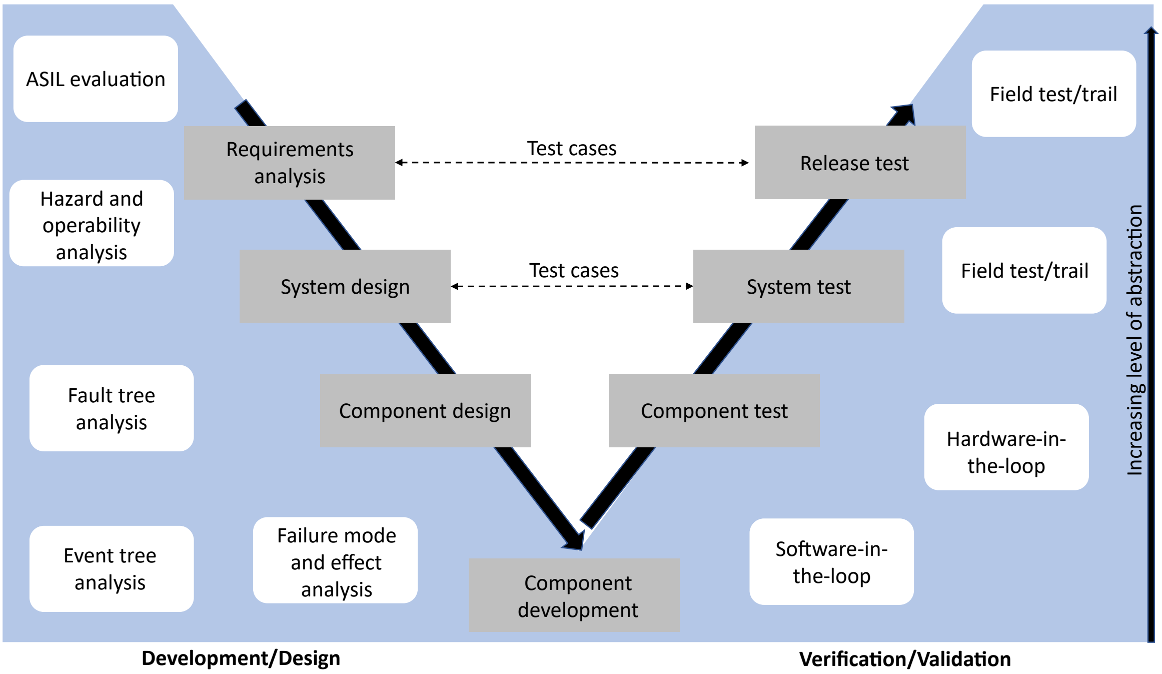


Figure 10. The current process of the development and proof of safety in a V-Model based on Weitzel et al. (2014), adapted from Wachenfeld & Winner (2016).

Early in the development process, tests are carried out in virtual test environments in previously defined test cases (e.g., software-in-the-loop). The further the development progresses, the more real components can be tested (e.g., hardware-in-the-loop, driver-in-the-loop, or vehicle-in-the-loop). However, simulation models are still needed to test the entire vehicle system. Ultimately, these simulation models always represent a simplification of reality and cannot fully reproduce it. For this reason, current systems are always tested with real vehicles, people, and environments at the end of product development. This procedure follows a track distance and statistical approach covering a required test distance under representative conditions in real traffic without accidents. (cf. Wachenfeld & Winner, 2016) For example, using real driving required a total of 36 million test kilometres before the Mercedes Benz E-Class (W212) was released (Daimer AG, 2009). According to Fach et al. (2010), the release of a current driver assistance system alone requires up to two million test kilometers. These examples show that even for current driver assistance systems, approval based on real driving in road traffic represents an economic challenge, growing given the increasing automation and variety of functions.

2.2 Special features of highly automated driving

As already described above, the focus for the approval of current systems is on the vehicle and especially its controllability by the driver. In the combined representation of the three-level model for human target-oriented behavior based on Rasmussen (1983) and the three-level hierarchy of the driving task based on Donges (2015) in Figure 11, this approval corresponds to the elements highlighted in blue. The vehicle and its behaviour in longitudinal and lateral directions are tested; however, not the behaviour or the abilities of the future driver, but only the possibility for the test driver to control the vehicle in the test cases by steering and acceleration interventions. Therefore, the blue box only slightly cuts the area that represents the driver. Additionally assumed, but not tested, is the reliability of the driver.

For HAD, the driver's skills are no longer required, and she/he no longer functions as a fallback level. Automation takes over the driving task, i.e., navigation, guidance, and stabilisation. The previously assumed and untested reliability of the driver disappears, is replaced by automation, and must be tested for the automated system. When proving the autonomous system, safety results only from the technically automated system and the vehicle (orange box of Figure 11). It can be seen that, compared to current systems, both the number of tasks in the form of more ODDs and the quality of tasks (in addition to stabilising tasks, there are also tasks in guidance, navigation, and independent monitoring) increase. (cf. Wachenfeld & Winner, 2016)

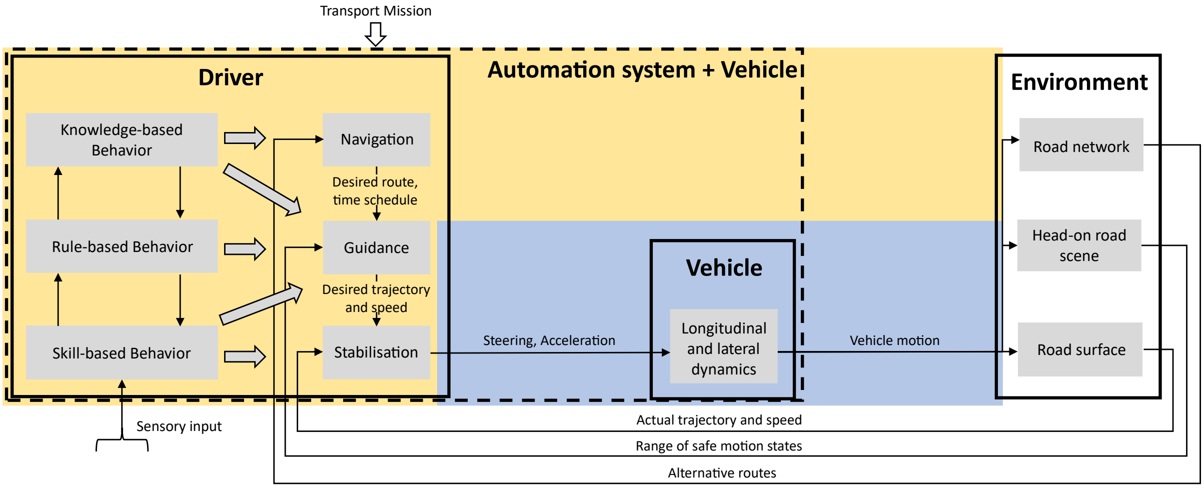


Figure 11. Three-level model for human target-oriented behavior based on Rasmussen (1983) and the three-level hierarchy of the driving task based on Donges (2015), adapted from Wachenfeld & Winner (2016).

Thus, function of HAD as a test object differs fundamentally from current road vehicles. The test concept consisting of test case generation and test execution, as mentioned previously, no longer applies directly to HAD, as described below. First, the test case generation is based on the assumption of the driver's driving capability. Whether a driver can control the test object is linked to the legally required driving licence. According to the Road Traffic Act (§ 2 Abs. 2 StVG), this driver’s license is only issued if the applicant, among other things:

- has attained a minimum age,
- is suitable for driving a motor vehicle (§ 2 Abs. 4 StVG), meaning to fulfill necessary physical and mental requirements as well as compliance with traffic regulations or criminal laws,
- has received training,
- and has passed theoretical and practical tests.

Due to these requirements, test case generation is limited to test cases, assuming that if the test driver is able to cope with these exemplary situations, any other driver with a driving license will also be able to cope with the other untested relevant situations in the field. These include situations in which the driver is driving manually and in which the driver monitors and overtakes direct control of the system if necessary. Thus, the test cases, in combination with the driver's license test, provide a metric that allows us to conclude regarding the safety of the driver-vehicle system. (Wachenfeld & Winner, 2016)

Due to the omission of the driver, the currently accepted metric is no longer applicable for HAD, and thus, the reduction of test cases is no longer permitted. Instead, the test case generation for HAD must cover the driving skills that the driver has previously brought into the driver-vehicle system. These capabilities are fulfilled by the human, among other things, by the fact that she/he:

- has experienced hundreds of thousands of kilometres as a road user,
- has experienced social behaviour as part of society,
- has learned cognitive skills
- and has trained sensorimotor skills.

This means that the driver does not only acquire her/his driving ability with the driving licence and its underlying tests, but it represents a complex process of different experiences and acquired knowledge in society since birth. Therefore, it is impossible to introduce a simple driving licence for automation as the performance of the driving task is a complex construct. In fact, we have to measure the entire performance of the automation system in terms of the three simplified information processing stages in relation to the driving task (Winner, 2016): perception, cognition, and action. Unfortunately, no valid metric or method has been proven so far. Thus, the commonly accepted metric and the reduction of test cases are invalid. (Wachenfeld & Winner, 2016)

Secondly, as mentioned in Subsection 2.1, in terms of test execution, real driving is currently the most crucial method for release due to validity and economic feasibility. In addition, a test driver is available for driving on public roads to drive or monitor the vehicle in every situation according to the task of the vehicle user. In terms of HAD, using the test driver would not be a real component of the vehicle as the driver does not have to supervise or intervene anymore. (cf. Wachenfeld & Winner, 2016) Therefore, in addition to the test case generation, the current test execution is not directly transferable to HAD.

2.3 Current approval trap

Despite the differences between partial automation and HAD regarding safety approval shown above, the current test concept could still be theoretically retained. Therefore, after theoretical and statistical considerations, Wachenfeld & Winner (2016) have shown what this means for HAD. They conclude that, for example, for a motorway pilot without reducing the test cases, 6.62 billion test kilometres without an accident occurrence would have to be completed on the public motorway to provide statistical proof of safety through real driving. This is economically and practically not feasible for HAD. Thus, the release becomes a great challenge or the so-called approval trap for HAD (Wachenfeld & Winner, 2016) because the release of serial production of prototypes cannot take place. Thus, a broad use of HAD in public road traffic is not achieved (see Figure 12). Additionally, other factors can increase the number of test kilometers. For example, a system variation would lead to the test kilometres being driven again. Winner (2016) details how the different parameters such as area of application, type of accident consequences, cause of an accident, and comparison vehicles affect the necessary kilometres determined.

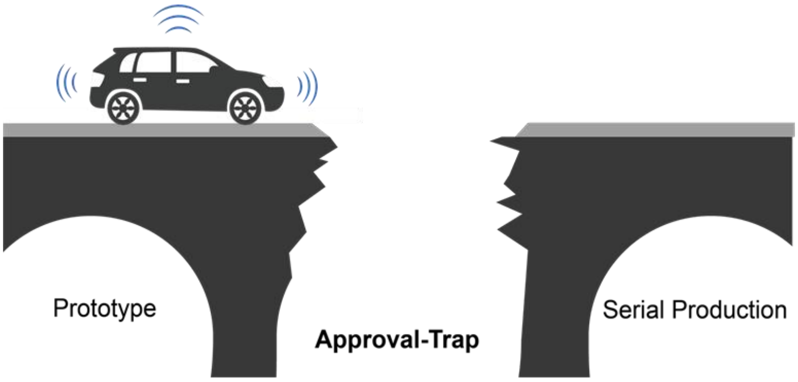


Figure 12. Visualisation of the approval trap.

Finally, the test dilemma can only be overcome by significantly shortening the required driving distance when still using the current approach or using a completely new approach and perspective. Thus, new test methods and approaches have to be developed. (Winner, 2016)

2.4 Common approaches to overcome the approval trap

In general, safety is commonly defined as the freedom from unacceptable risks and dangers in the change of location of persons or material assets (traffic objects) that are transported, for example, utilising transportation from A to B. This includes the transport infrastructure and transport organisation. (Schnieder & Schnieder, 2013, p. 74) A basic distinction has to be made between two points of view in terms of safety (Schnieder & Schnieder, 2013, p. 67):

- protection of the environment from system impacts, which is referred to as *safety*

- protecting a system from external influences, which is called *security*

Currently, the automotive industry addresses safety mainly by the *safety of the intended functionality* (SOTIF; ISO 21448:2022), *functional safety* (ISO 26262:2018), and *security* (e.g., cyber attacks) based on the SAE J3061 (2016). This thesis only deals with the aspect of safety and not security. SOTIF ensures an intended function by preventing hazards due to functional deficiencies in the absence of technical system failures (ISO 21448:2022). The intended function is related to the *object and event detection and response* (OEDR), including monitoring the driving environment and executing an appropriate response to objects and events in a specific ODD. In contrast, functional safety ensures that the intended function does not induce further hazards caused by technical malfunctions due to random or systematic faults in the system's hardware or software (ISO 26262:2018).

Two primary methodologies have to be distinguished for designing and assessing the functional safety- and SOTIF-related capabilities of automated vehicles: *risk identification and evaluation* methods which serve to deliver system requirements and system design recommendations, ensuring a system is as safe as possible and *safety validation* methods to prove that the safety requirements are actually fulfilled (see Figure 10). Both types of methodologies contribute to the safety of automated vehicles. However, the approval trap primarily arises due to an unsolved safety approval belonging to safety validation methods focused in the following. Research, e.g., in the form of frameworks and toolchains such as developed in PEGASUS (German Aerospace Center [DLR]) and VVM (VVM consortium), standardisation like the UL 4600 (Koopmann, 2022), and a code of practice created in L3Pilot (Cao et al., 2022), is being carried out around a combination of various existing methods enabling the soundest evidence concerning safety in order to overcome the approval issue.

In the following, based on Riedmaier et al. (2020) and Junietz et al. (2018), a brief overview of frequently suggested or used validation methods is given, which can be differentiated into macroscopic and microscopic assessment, i.e., a statistical statement about the overall system effect or evaluations of individual scenarios, respectively, as defined by (Junietz, 2019) (see Figure 13):

- <u>Formal verification:</u> formal proof on an abstract mathematical model of the system depending on assumptions and formalised rules that have to hold true. Currently, this approach lacks scalability for complex systems. (e.g., Shalev-Shwartz et al., 2017)

- Performance metric: the basic idea is to define a performance metric, i.e., a quantitative measurement of the extent of a property a system has. Therefore, Winner (2016) suggests defining a metric that compares the general driving task performance between drivers and automation, distinguishing the perception-, cognition-, and action performance. Currently, how to construct and operationalise such a metric is unknown, especially due to the cognitive processes. (Winner, 2016)
- Traffic simulation-based: shifting the safety assessment from the real world into traffic simulation of the whole road network with hundreds of road users (so-called agents) increases the validation efficiency, assuming that the simulation model is valid to represent the real world but is quite challenging. (e.g., Kitajima et al., 2019; Roesener et al., 2018)
- Shadow mode: running an automated driving function passively in the background of a manually driven vehicle. The function is provided with sensory inputs but cannot access the vehicle actuators. Ultimately, the function's decisions are evaluated. One major disadvantage is that the behavior of the other interacting road users is unrealistic, as the automation can decide differently than the driver. Thus, other road users would have acted differently than the actual behavior, resulting in other interactions. (Wang & Winner, 2019)
- Staged introduction and decomposition: the concept is to limit the ODD and thus reduce the validation effort by real-world testing to an economically feasible way and then gradually increase the ODD and only validate the new ODD-related functions following the principle of functional decomposition. (Amersbach & Winner, 2019; Wachenfeld, 2017)
- Real-world testing: as represented by the approval-trap, real-world testing is not feasible. However, the extreme value theory can be used to apply surrogate metrics like time-to-collision (TTC), which can be extrapolated to predict rare events such as an accident. The assumption is that critical events, i.e., incidents, happen more frequently, reducing required data collection milage and that the incidents' frequency can point to the likelihood of an accident. (e.g., Asljung et al., 2017)

- Function-based: specific system functions are tested in a few fixed and limited tests on a test track or in a simulation. A pre-requirement is the definition of system functionalities, which is impossible for every possible situation due to the open parameter space for HAD comprising an inherent uncertainty. (e.g., UN ECE R131, 2013)
- Scenario-based: the aim is to reduce the test scope by only addressing relevant scenarios based on critical events. However, the challenge is defining a representative but efficient set of scenarios. Two different selection processes that follow the ground scientific reasoning principles of induction and deduction can be differentiated: testing-based to cover the parameter space with finite test cases or falsification-based by focusing on challenging edge cases to find counterexamples. Even if all test cases are successfully passed, and no counterexamples can be found, safety still cannot be guaranteed entirely because both procedures are microscopic assessments, and conclusions cannot be drawn for the entire system. Thus, an additional macroscopic assessment method, e.g., formal verification, can be used. (e.g., Riedmaier et al., 2020)

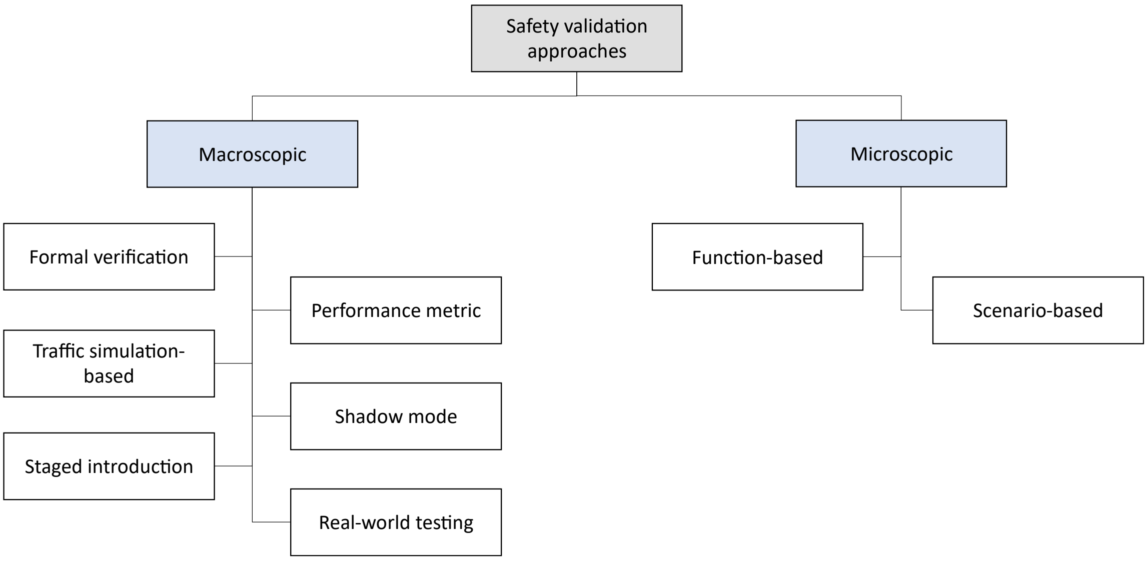


Figure 13. Overview of safety validation approaches frequently suggested or used in automotive, adapted by Riedmaier et al. (2020).

According to Junietz et al. (2018), the mentioned validation methods differ in three basic dimensions: the object under test (OuT), meaning what is tested; the stimulus used to trigger a reaction of the OuT, meaning how it is tested; and the assessment criterion defining how is assessed. For example, the methods have differences in the abstraction level in which parts of the system are evaluated, in the accuracy level to represent the real world and entities, and in the direct or indirect (surrogate) measurement of the interested aspect.

Nevertheless, every approach faces disadvantages, which is why a combination of different approaches is recommended to decrease the residual risk to an acceptable level (Junietz et al., 2018), e.g., the integration of the scenario-based approach with the formal verification seems to be promising (Riedmaier et al., 2018).

2.5 The naive fallacy and a new perspective

Traditionally, road safety issues are addressed by adopting a deterministic and reductionist approach which involves decomposing the system into its component parts, examining the parts and improving their performance in isolation, and reintroducing them back into the system (Read et al., 2017). This approach produced successful measures resulting in positive outcomes (see Figure 8). However, only parts of the road system were improved, not considering the inherent complexity of the system or the full range of factors shaping the behaviour (e.g., Cornelissen et al., 2015; Larsson et al., 2010; Salmon et al., 2012; Salmon & Lenne, 2015) to understand how these parts interact together and how the entire system works (Read et al., 2017). The safety trends are plateauing as the traditional approach reaches its effectiveness limit (Salmon et al., 2017). Therefore, a shift to address road safety issues by a systems thinking approach is needed (e.g., Hughes et al., 2015; Larsson et al., 2010) to understand that supposed causes like driver errors, in fact, mostly represent the effects of system-wide issues as symptoms, rather than the primary cause of accidents (Read et al., 2017).

These two contrasting safety approaches can also be related to the historical development of the scientific study of safety, pointing out two fundamental concepts concerning safety: safety-I and safety-II (Hollnagel, 2014). Safety-I is described as a situation where as few things as possible go wrong. The common assumptions are (Hollnagel, 2019b):

- the system can be decomposed into meaningful elements
- the function of each element is bimodal (true/false, work/fail); success and failure are seen as separate states due to system functioning (work-as-imagined) and malfunctioning (non-compliance error)

- the failure probability of elements can be analysed individually
- the order of events is deterministic
- systems are well-designed, and designers have anticipated every contingency and thus provided complete, correct, and appropriate response measures
- operators behave as expected and have been trained to

This point of view assumes that adverse outcomes are caused by technical, human, or organisational failures and malfunctions, which must be eliminated or prevented by proper barriers. The human is seen as a liability, and performance variability should be prevented. This perspective evolved in the 1920s when systems were loosely coupled, linear, and stable, and system functions were easy to understand and well-understood (Hollnagel, 2014). However, most current systems are tightly coupled, increasingly non-linear, less stable, and system functions are hard to understand due to their complexity. Those systems are intractable, and outcomes cannot be totally controlled or predicted. Thus, the perspective changed to safety-II. Safety-II is seen as a situation where as many things as possible go right. The purpose is to understand how things usually go right to explain how things rarely go wrong. This perspective regards humans as an inevitable resource for system flexibility and resilience, whereby performance variability should be monitored and managed. The basic assumptions are (Hollnagel, 2019b):

- systems cannot be understood by decomposing them
- functions are not bimodal; in fact, performance is always variable
- this performance variability is a source of success as well as failure
- the functions must be flexible to fit the conditions, which is noticeable as work-as-done instead of work-as-imagined
- most events emerge due to complex interactions among system elements and their interdependent performance variabilities and rarely result from clear cause-effect relationships

As we can see, the analyzed system's characteristics and the analysis's granularity level are essential when choosing approaches or methods to manage safety issues. Salmon et al. (2012) concluded that the road system, which connects technical, psychological, and social elements to transport people and goods from one place to another, is of a socio-technical nature. Additionally, these authors demonstrated that the road traffic system is complex based on the prerequisite properties of complexity presented by Dekker et al. (2011). Consequently, the road transport system is a complex STS. It thus could be embedded in the systemic quadrant of the system interaction-coupling matrix adapted from Perrow (1984) when combined with the accident analysis methods classification by Wienen et al. (2017) (see Figure 14). This assignment implies that systemic methods are best suited to represent safety assessments in road traffic in general. It should be noted that the assignment of systems as well as model categories is rather notional than such distinct in reality depending on the changing operating conditions of a respective system, which is why the respective placements are debatable as no reliable and valid measurement of the dimensions exists (Perrow, 1984). For example, Hollnagel & Speziali (2008) applied the assignment slightly differently, positioning epidemiological methods differently.

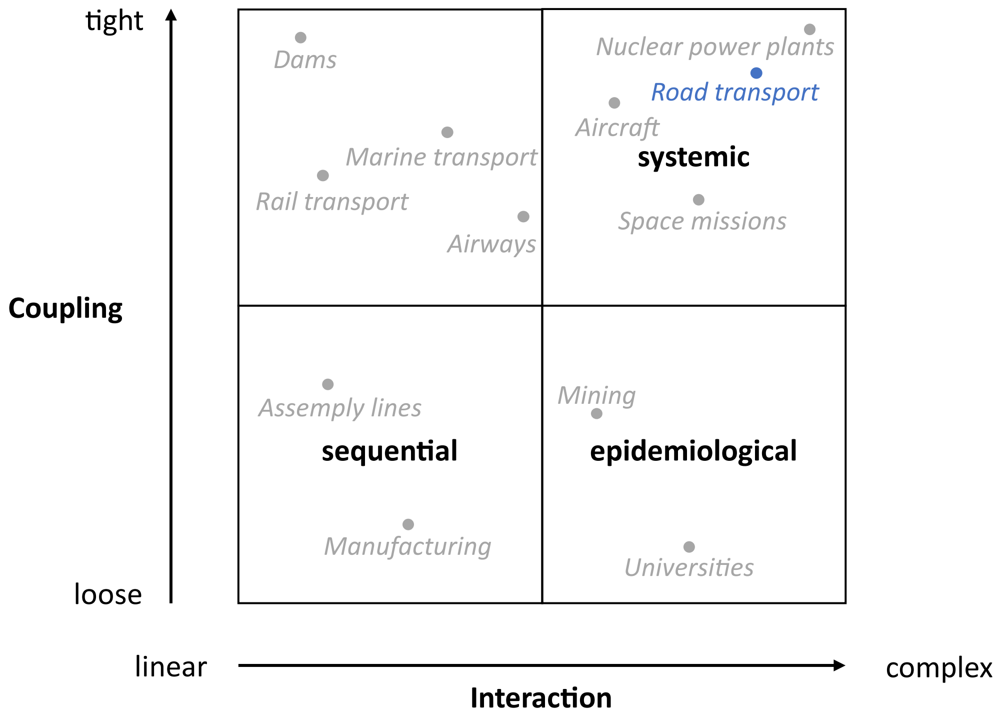


Figure 14. System interaction-coupling matrix combined with accident analysis methods classification and assignment of several systems, especially the road system, adapted from Perrow (1984), Wienen et al. (2017, p. 22), and Grabbe et al. (2020b).

A critical perspective on the two safety views in relation to automated driving and road safety unveils one interesting fact: the point of view of safety-I supports the argumentation of the substitution of the driver by automation, and on the opposite, the point of view of safety-II argues in favour that the driver is still necessary for at least some situations due to system flexibility and that current automated systems probably are not able to cope with this flexibility (cf. De Winter and Hancock, 2015) in any situation. This indicates that the common motivation of full automation and the associated goal of increased traffic safety is essentially safety-I driven, and the safety-II perspective is almost neglected. In fact, this is also reflected in the aforementioned definition of traffic safety in Section 2.4, which is largely safety-I oriented.

In particular, the common approaches to overcome the approval trap, presented in Section 2.4, have in common that they still follow the traditional reductionist approach and a safety-I perspective because they only shift and optimise rather than solve the problem. This is consistent with the insight by Salmon et al. (2012) that the whole approach to understanding and enhancing behaviour and safety in road transport is entrenched within the reductionist philosophy as the mainstream reasoning. This also aligns with Zhang et al. (2021), who recognised that functional safety by ISO 262622:2018 and SOTIF by ISO 21448:2022 are logically rooted in a Newtonian mechanistic world and do not include systemic techniques. For example, the underlying processes are built upon the assumption that the system is completely specifiable and methods predominantly used for risk assessment like the failure mode and effects analysis (FMEA, Kirwan and Ainsworth, 1992) and failure or event tree analysis (FTA, Watson, 1961) rely upon the reductionist causality credo based on the event chain model of failure development (Leveson, 2011; Thomas et al., 2015). Even though a systemic method by the system-theoretic process analysis (STPA, Leveson & Thomas, 2018) based on systems-theoretic accident model and processes (STAMP, Leveson, 2004) was added as a tool to the ISO 26262:2018 in recent years, this method still mainly follows a safety-I thinking as it is based on control theory and not complexity theory (Grabbe et al., 2020b). Ultimately, Zhang et al. (2021) criticise that the current methodologies used in automotive safety evaluations lack the understanding of human-automation interactions claiming to promote system-thinking tools from the human factors discipline which acknowledge that road traffic is a complex STS that shapes the behaviour of drivers and other road users (Lintern, 2020). Even the standards by the International Organisation for Standardisation (ISO) related to the area of risk analysis and risk management in general do not fit to identify risks arising from complex interactions and emergent behaviour (Björnsdottir et al., 2022).

It can be argued that all approaches to overcome the approval trap commonly following the safety-I perspective only evolutionarily optimise the problem but do not completely solve the problem in the sense of a revolution because they are rooted in a naive fallacy. This means that potential solutions addressing the approval trap follow the same kind of thinking that already created the approval trap. In other words, the problem has to be thought of differently by finding the right problem rather than creating the right solutions for the wrong problem. The naive fallacy can be seen as an enhancement of the oversimplification fallacy mentioned in Section 1.4. This phenomenon is also in line with Hollnagel's (2019a, last slide) general statement regarding safety management that "it is an unavoidable dilemma that we inadvertently create the challenges of tomorrow by trying to solve the problems of today with the mindset (models, theories & methods) of yesterday".

Here comes the safety-II perspective and the systemic approach into play. Their application seems urgently required (Grabbe et al., 2020a, b; Papadimitriou et al., 2022). Apparently, there is no "one-size-fits-all" solution to safety, especially for complex and dynamic STSs. Thus, overall, we need combinations of different views, approaches, and measures, including the use of safety-I and safety-II in a complementary manner. However, a significant perspective, that is a complexity-oriented holistic approach based on resilience engineering (RE) (Hollnagel et al., 2006) which considers interactions, processes, and patterns within a complex system that form the adaptive capacity to be resilient, is currently lacking and inevitable as a fundamental basis for the safety assessment of automated driving. This potentially helps to reveal hidden risks or safety blind spots of automated driving in relation to the overall traffic system performance. In particular, this approach could be the answer to the outstanding issue of what can be opposed to the previous thinking in terms of the bimodality "right or wrong" (Winner, 2016) to overcome the approval-trap and the statistical approach which is based on track distance that currently focuses only on the counting of rare, adverse events.

3 Systems Thinking versus Reductionism

> "94% of problems in business are systems driven by only 6% are people driven."
>
> – W. Edwards Deming –

As we have seen above, the safety assessment of automated driving can benefit from a systemic approach based on RE. Therefore, this chapter provides a solid foundation about the historical development of the scientific study of safety due to changes in the nature of systems, illustrating the transition from newtonian reductionism towards systems thinking and today's RE.

3.1 Traditional risk and safety management: Newtonian reductionism

Newtonian logic or reductionism (Dekker, 2011) allows describing any phenomenon by decomposing systems down to their component parts and their analysis, assuming that the overall system behaviour can be fully understood if the individual components can be understood taken separately (Walker et al., 2010). This follows reductionism and a mechanistic functioning of the world (Heylighen, 1989) which is strongly rooted in Western culture (Hollnagel, 2012a; Sacks et al., 2014): see, for example, Leucippus of Miletus (c.480–c.420 BC) – a greek physicist and philosopher who developed the atomic theory – who said: "nothing happens in vain, but everything from reason and by necessity" (cf. Taylor, 1999). This is similar to Newton's third law of motion, saying "action equals reaction". For safety and risk management, this reasoning implies thinking in bimodality which is about the functioning and non-functioning of a system as a result of functioning and non-functioning states of its individual parts, claiming that acceptable and unacceptable outcomes happen due to distinct modes of functioning (Braithwaite et al., 2015). Accidents occur as a result of one or more failures in various components, such as machine malfunctions, human errors, or non-compliance with procedures. The Newtonian approach to explaining accidents adheres to a causality credo, examining linear cause-and-effect relationships that underlie an accident. It assumes that every phenomenon arises from deterministic and identifiable causes, leading to definitive and identifiable effects, such as A causing B. This framework allows for identifying a causal chain that can be traced back to a root cause responsible for an adverse event. Consequently, in reactive risk assessments like accident analyses, the critical task is to identify these root causes and eliminate or prevent them to avoid similar adverse events occurring in the future. Likewise, in proactive risk assessments, it becomes possible to predict outcomes with absolute certainty and accuracy by understanding the system's initial conditions, enabling anticipation of future adverse events with harmful consequences.

3.2 The three ages of the scientific study of safety

Reductionism's legacy became pervasive in traditional safety science in the sense that newtonian reasoning is equated with the scientific study of safety (Patriarca, 2017). It can be seen as the underlying foundation for initial activities concerning safety and risk in the early 1900s, summarised as scientific management by Taylor (1911), including, e.g., the systematical recording of accidents and providing workers with protection from equipment in order to improve the working places and their environment. As the reductionist approach may have been adequate for the systems that existed at the time, i.e., simple and closed systems (Dekker, 2011), it has become more and more inadequate over time until today for present systems due to changes in the characteristics of the systems (Leveson, 2011). This is reflected in the historical development of the scientific study of safety, where many different models, methods, and paradigms evolved, as Dekker's (2019) comprehensive overview of the theoretical foundations of safety science has shown.

According to Hale & Hovden (1998), three common "ages" can be separated where each theory's, method's, or model's perspective on safety and risk corresponds to one more or less. Notably, these different perspectives supplement, rather than substitute, each other (Hovden et al., 2010). The first age of technology covers the period from the early 1900s up to the Second World War, involving technical measures to prevent the occurrence of technical and mechanical breakdowns. Then, the second age of human factors integrated the impact of human performance, mainly considered as a limitation in safety and risk management. Lastly, up from the 1990s, the age of safety management, coping with complex STSs, arose acknowledging systems' complexity by shifting away from an exclusive focus on individual error towards the role of multiple actors on all levels of a system and understanding complex system success and failure as emergent properties of interdependent system elements shaped by socio-technical factors. These ages can also be partly mapped to the three types of accident models proposed by Hollnagel (2002): sequential, epidemiological, and systemic analysis techniques. As shown in Figure 15, sequential techniques are by far the oldest, while systemic approaches only appeared 20 years ago. Another classification option for the accident analysis and risk assessment methods is to break these down into technical, human factors, organisational and systemic methods (Eurocontrol, 2009), see Figure 15.

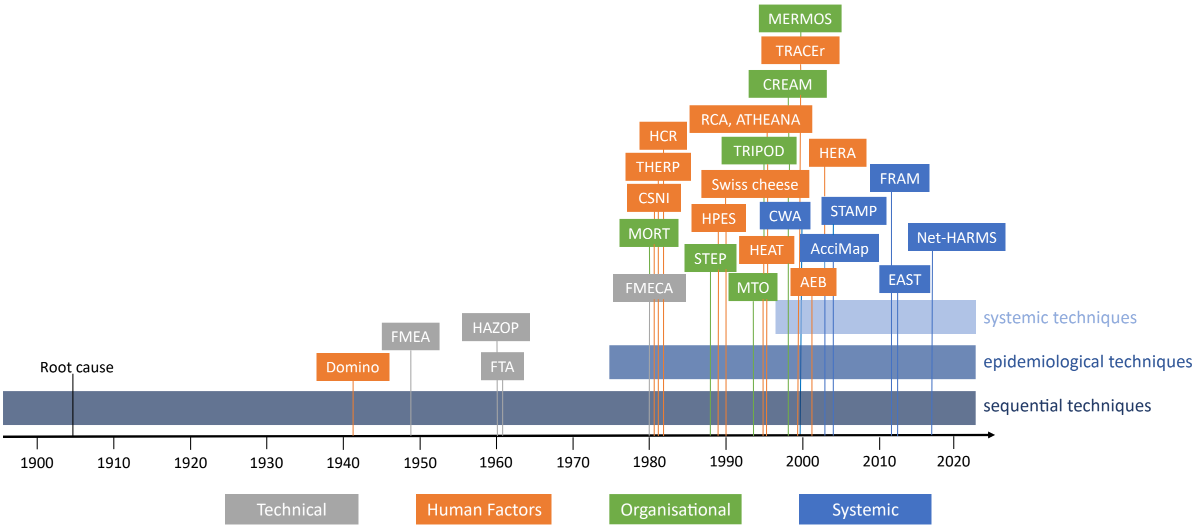


Figure 15. Overview of the development of common accident analysis techniques and important accident analysis and risk assessment methods adapted from Underwood (2013, pp. 18-19, 27) and Eurocontrol (2009), adapted from Grabbe et al. (2020b).

The basic principles of the sequential, epidemiological, and systemic analysis methods are shown below. Sequential accident models describe the accident as resulting from a chain of discrete events occurring in a particular time sequence. Here, losses are caused by technical failures or human error, assuming that the cause-and-effect relationship is linear and deterministic (Qureshi, 2007). These methods follow the Domino Theory introduced by Heinrich (1941), including, among others, the FTA and FMEA. In the mid-1980s, further factors or conditions and explanations were needed to understand the disaster at the Chernobyl nuclear power plant or the loss of the space shuttle Challenger. Thus, the focus changed from human factor to organisation leading to epidemiological techniques (Hollnagel, 2012a). New concepts and theories, such as Reason's (1990) swiss cheese model or the cognitive reliability and error analysis method (CREAM) of Hollnagel (1998), were thus developed to explain accidents as a result of a combination of different interacting, active and latent factors on different hierarchical levels (Qureshi, 2007). This improved the understanding of accidents regarding complexity. However, the focus was primarily on the “sharp end” factors (Dallat et al., 2017), and the causality is still linear with links between states that are loosely coupled, which does not adequately represent the dynamics of a system (Hollnagel, 2004). Thus, systemic models arose, seeing the accident process as a complex and interwoven event that cannot be broken down into individual parts (Wienen et al., 2017) and rather analysing interactions within the whole system. New accident models had to be developed based on system theory (Leveson, 2004). The most widely-used systemic models are Leveson's (2004) STAMP and Hollnagel's (2004) first proposal of the functional resonance accident model (FRAM) and later adapted to the functional resonance analysis method (Hollnagel, 2012a).

As the historical development of safety management illustrates, humans have an intrinsic desire to understand the world around them to satisfy their need for certainty and to feel in control, which is accomplished by using theories, models, principles, and methods which must be in accordance with reality. These concepts have changed over time as the world is not constant and stable but rather dynamic. At first, this progress was slow, not posing a problem, but since the 1950s, the changes have been rapidly speeding up, so humans can no longer keep pace. This is because we build larger and larger systems following the credo for faster, better, and cheaper systems, resulting in complex STSs where the assumptions of reductionism or safety-I are increasingly less representative of understanding them. (Hollnagel, 2012a)

This led to systems theory trying to understand the behaviour of STSs and their inherent complexity. Before introducing the concept of systems theory, further attention has to be devoted to STSs and complexity, which represent the core features of systems thinking.

3.3 Features of socio-technical systems and complexity

Socio-technical systems theory appeared in the 1950s when Trist & Bamforth (1951) argued to focus on optimising both technical work processes and the social systems operating within the work environment to improve organisational performance. An STS comprises interconnected social and technical elements that mutually influence each other, either directly or indirectly, to sustain their functioning and ensure the system's continued existence to achieve its objectives. In addition, these elements are affected by environmental conditions, which they impact in return (Pasmore et al., 1982). Figure 16 schematically depicts the features of an STS. The social system consists of human beings who work in the organisation at different levels and the relationships among them. The technical system comprises technological artefacts designed to transform inputs into outputs in a task-based manner. The interdependencies are as follows: occupational roles indicate the relationship between people and tasks; the organisation defines work processes of how technology is used to produce outputs, while specifying procedures of how a task should be performed; artefacts are related to the skills and capabilities of people. Moreover, the social system affects the technical sub-system based on allocated resources, determined goals, cognitive capabilities, and work constraints. The technical system influences the social sub-system through technological functionalities, capabilities, and constraints. Both systems are constrained by the environment, e.g., through the operation conditions, or influence the environment. It should be noted that STSs progressively include interconnected cyber-technical artefacts, thus becoming cyber-socio-technical systems (CSTSs) (Patriarca et al., 2021). However, in this thesis, the term STS is used more broadly, where artefacts can be both physical-technical and cyber-technical.

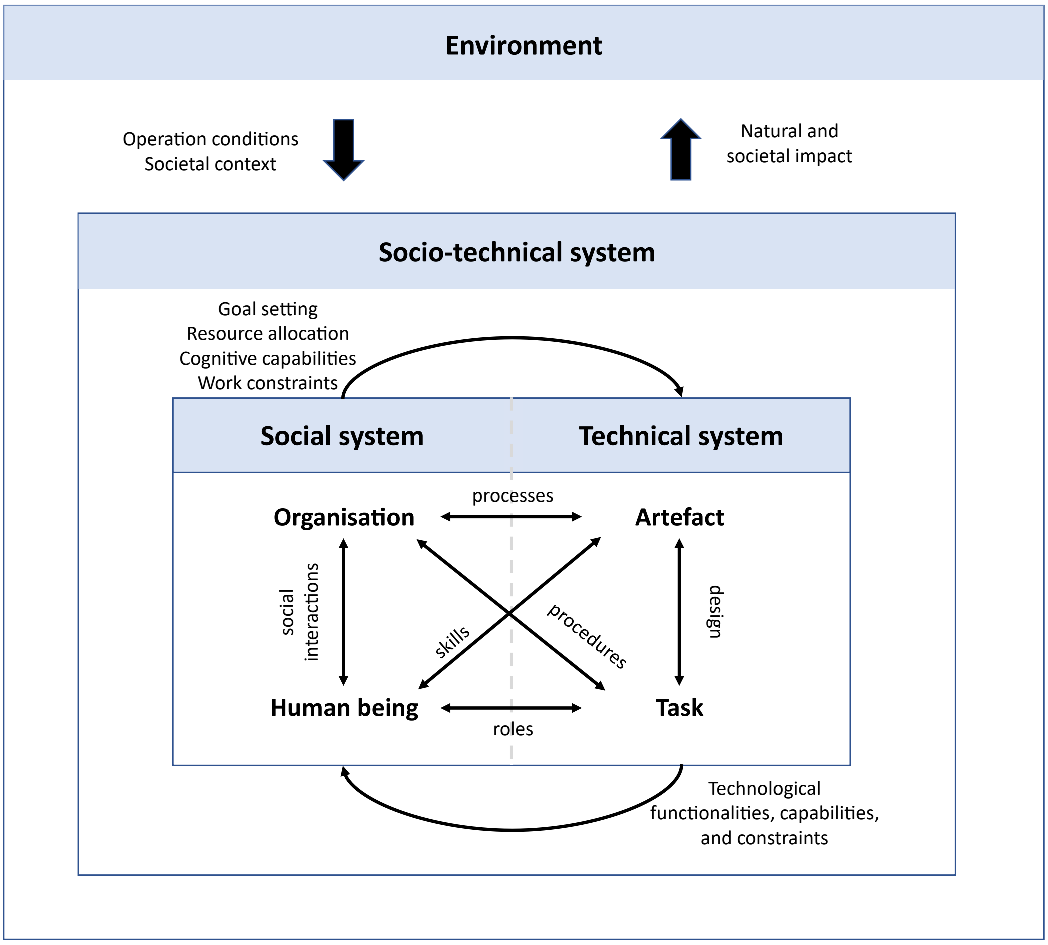


Figure 16. Schematic representation of a socio-technical system, adapted from Patriarca (2017) based on Bostrom & Heinen (1977) and Di Maio (2014).

The interaction of these social and technical elements comprises partly linear cause-effect relationships and partly non-linear and complex ones creating emergent behaviour leading to successful or unsuccessful system performance (Walker et al., 2009). A key tenet for safe and efficient performance in STSs is the adaptive capacity (Read et al., 2017) which is primarily achieved by joint optimisation (Emery, 1972) in the sense of coagency in a joint cognitive system (JCS) (Hollnagel & Woods, 2005) taking into account the functional entanglement of the two sub-systems as opposed to the isolated optimisation of technical and social elements. Underpinned by systems theory, STSs align with open-systems principles formulated by Skyttner (2001) mainly based on von Bertalanffy's (1950) work regarding the general systems theory. According to Underwood (2013), the open-systems principles can be differentiated into three groups: system structure, system component relationships, and system behaviour (see Figure 17).

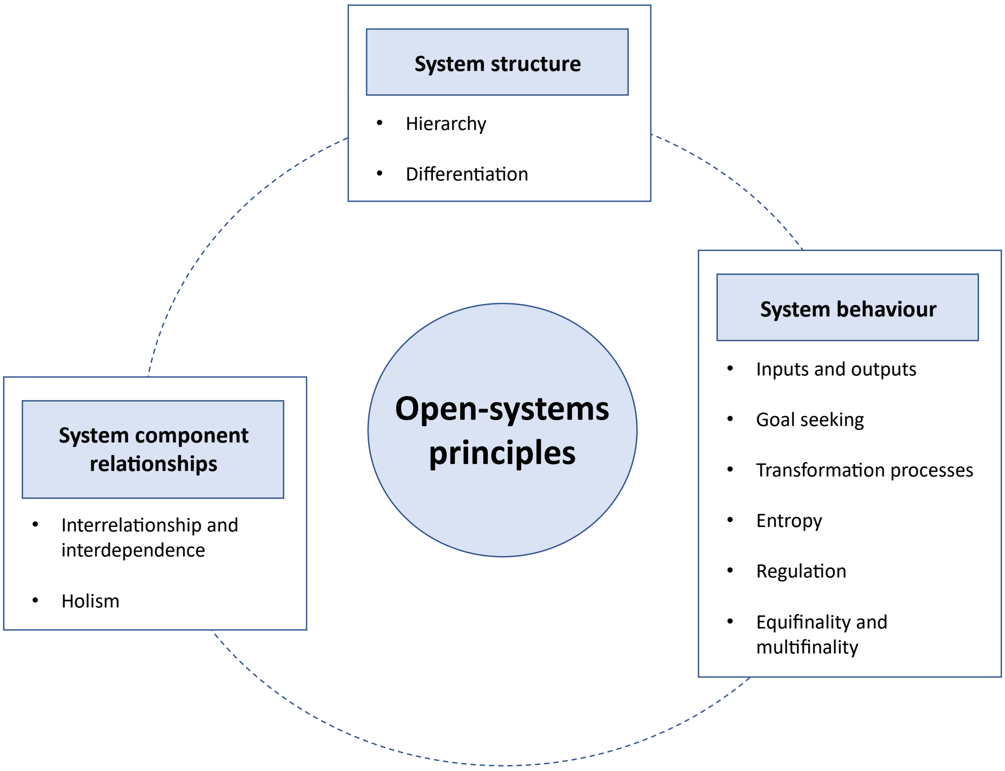


Figure 17. Overview of open-systems principles by Skyttner (2001), adapted from Underwood (2013).

Systems comprise sub-systems nested within one another, also called system of systems (Von Bertalanffy, 1968), following a *hierarchical* structure. The sub-systems are formed to perform specific functions, known as *differentiation*. In order to specify a system's hierarchy, the boundary of a system has to be determined, i.e., distinguishing between what is part of the system and part of the environment (Vicente, 1999).

System components are *interrelated* and *interdependent*, meaning that one component influences the other parts or is affected by them directly or indirectly. Therefore, the interaction of system components produces emergent, rather than resultant, properties. Hence, the whole is more than the sum of its parts. Consequently, a system must be studied *holistically* and not analysing the parts in isolation. *Inputs* are received from the environment and *transformed* into *outputs* transferred to the environment to achieve the system's goals. The system behaves as *goal-seeking* as the interactions result in some goals, a final state, or some equilibrium to be approached. To obtain these desired goals, the interrelated components must be regulated through control and feedback loops in an adaptive way where the transformation processes are adjusted to fit the input-output relation. The system's level of *entropy*, i.e., the disorder or randomness in a system, tends to increase without intervention. These former principles result in dynamic system behaviour that can achieve a goal from various initial starting conditions (*equifinality*), or systems can produce a range of outputs or different and mutually exclusive objectives from the same initial starting point (*multifinality*).

A particular challenge of STSs is to comprehend the inherent complexity arising from the interactions between multiple system artefacts and social agents distributed in time and space while engaged together to ensure the system's goals (Harvey & Stanton, 2014). Following a broader perspective, STSs can be seen as a special case of a complex adaptive system in which the structural and dynamic properties adaptively adjust in response to internal and external perturbations (Miller & Page, 2007). Here, complexity theory needs to be applied to understand how STSs function in order to develop design changes that might improve their functioning (Pavard & Dugdale, 2006). In general, the word complex comes from the Latin complexus which means “what is woven together”. Research on complexity has in common to be interested in systems where multiple interacting and intertwined elements create hardly-identifiable patterns to which they are able to adapt or react, causing non-linear and unpredictable propagations through the system (Arthur, 1999). These complex interactions may lead to dynamic events characterised by processes that vary asymmetrically and irregularly with non-trivial functioning principles rather than controlled by simple cause-effect relationships (Feltovich et al., 2004).

In an epistemological view, an important distinction has to be made between “complicated” and “complex” systems (Dekker et al., 2013). Both systems have in common to comprise a multitude of interacting components, but that is where their commonality ends (Cilliers, 1998; Heylighen et al., 2007). According to Dekker et al. (2013), complicated systems are ultimately knowably affording a complete, exhaustive description by a set of rules that can fully capture their workings in a linear way. This makes complicated systems predictable and controllable, similar to a machine. The whole is equal to the sum of its parts. For example, a jet airliner or a passenger car are complicated systems. They contain thousands of mechanical parts, and understanding how they work might be difficult for a single person; nevertheless, they are understandable and describable in principle (Dekker et al., 2011). In contrast, complex systems are neither fully knowable with the impossibility of attaining a complete, exhaustive description (Cilliers, 2002) nor a set of rules can be defined that can fully capture their functioning due to intractability (Page, 2008). The whole is more than the sum of its parts. Complex systems are open systems changing in interaction with their environment, where complexity emerges from a network of local interactions. This means that each component has a limited horizon concerning the consequences of their local behaviour up to the level of global system behaviour, resulting in the phenomenon that any agent's action controls very little but influences almost everything (Dekker et al., 2013). Hence, looking at micro-macro connections, local behaviour can produce global effects which are unpredictable at a local level (Dekker et al., 2008). Thus, more than one description of complex systems is always possible and even necessary due to dynamic, unpredictable, and multidimensional problems - no intelligent designer or governor has overall control over any non-trivial complex STS (Dekker et al., 2013). Returning to the above example of the passenger car, identified as a complicated system, and deciding to perform some maintenance on that vehicle, we create a new system, " vehicle maintenance," which becomes complex. The reason is that technical elements are interrelated to human, social, and organisational parts (e.g., policies, procedures, culture), where the system is opened to influences beyond engineering specifications and reliability predictions (Dekker et al., 2011).

Adapted from Goldratt (2008), the differences between complicated and complex can be schematically described as the following (see Figure 18): system A, whose fixed and actual couplings make it a complicated system (under the hypothesis that no more hidden links are present among the system's components); and system B, whose complexity is defined by multiple degrees of freedom illustrated through both potential couplings between system's components and hidden couplings between known system's components or unknown components in the system or environment. Even if the formerly mentioned hypothesis is not verified, system A is less complex than system B because some degrees of freedom of system A are constrained. In system A, outcomes are resultant due to clear causality where causes are as real as the effect. Instead, in system B, outcomes are emergent due to transient combinations of conditions only present at a particular point in space and time, making causes elusive rather than real (Hollnagel, 2012a).

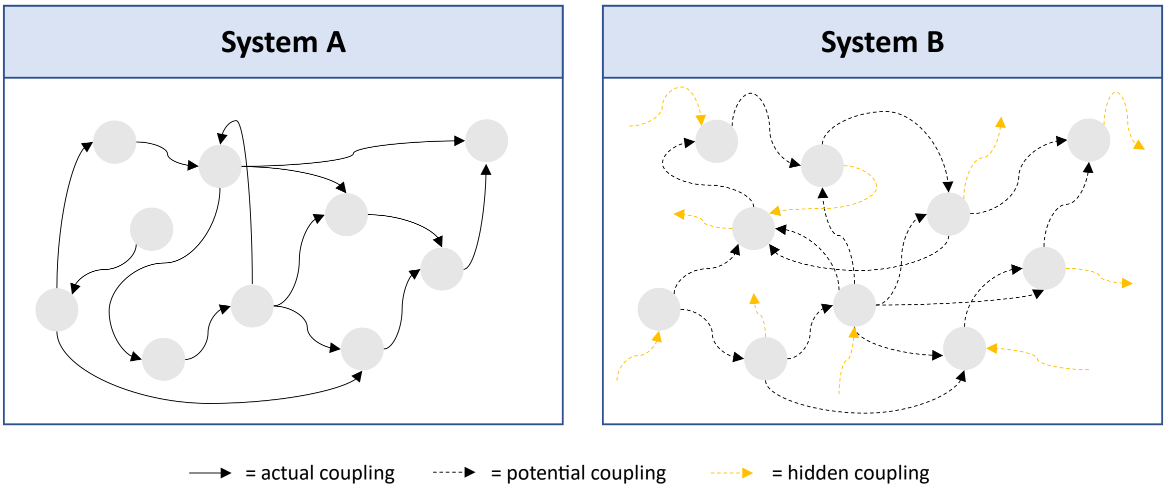


Figure 18. Complicated system A versus complex system B, adapted from Goldratt (2008).

Complexity is difficult to define (Cilliers, 1998), and various authors have outlined characteristics present in complex systems (Cilliers, 1998; Holland, 2014; Skyttner, 2001; Von Bertalanffy, 1968). However, a complex system is more defined by its relationships than by its constituent parts because many components are connected by non-obvious relationships, which makes up complexity (Hollnagel, 2012b). Furthermore, Rasmussen (1979) acknowledges that complexity is not considered a thing per se, rather, it is a situation to be investigated. The implication for safety management is thus that safety in complex systems is not a permanent property of what a system has but rather what it does, i.e., a dynamic non-event (Hollnagel, 2014). Nevertheless, it is possible to define a set – not necessarily complete and unique – of common characteristics of the complexity of interest for driving in the road system (Salmon et al., 2012) analysed in this thesis (Cilliers, 1998):

- complex systems are open systems in that they are open to influences from the environment in which they operate and also influence the environment in return

- each system’s component is ignorant of the system's behavior as a whole and does not comprehend their actions’ effects on the behaviour of the overall system
- it is the system that is complex rather than the components themselves, meaning that the system as a whole exhibits emergent properties that none of the components have, e.g., in a simplified view - only the combination of a vehicle, driver, and road can drive from A to B but not the parts itself
- components must continuously make inputs to keep the system functioning, which is why complex systems are dynamic and do not operate in a state of equilibrium
- complex systems have a history or path dependence manifesting in the influence of previous decisions and actions on the present time
- interactions in complex systems have recurrent loops, meaning that the effect of activities can feed back onto itself, directly or indirectly, which results in positive (amplifying) or negative (dampening) feedback loops
- interactions within complex systems are non-linear, characterised by an asymmetry between input and output, which is why small events can result in large effects or vice versa

3.4 Modern risk and safety management: Systems theory and resilience engineering

The term systems thinking describes a way of thinking about the reality based on systems theory aiming to understand and improve the performance and safety of systems and their surrounding environment humans are living in (Kim, 1999). It is a philosophy responding to the limitation of a reductionist and mechanistic ideology to comprehend social, socio-technical, and biological phenomena (Skyttner, 2005) by acknowledging the features of STSs and their inherent complexity, as mentioned above. In a nutshell, systems thinking looks at relationships, interactions, processes, patterns, dynamics, context, and the whole rather than isolated and decomposed parts, structures, outcomes, and statics. This requires multiple and different perspectives rather than a single one. The systems theory approach started to emerge from the 1950s onwards, by contributions of several researchers, for example, von Bertalanffy (1965), Wiener (1965), Ackoff (1971) or Checkland (1981), to deal with the increased complexity of the systems being built after world war II. However, von Bertalanffy is credited as the founder and pioneer of the entire systems thinking approach of what he called general systems theory (Leveson & Thomas, 2018). Some unique aspects of systems theory are:

- The focus is on the whole, i.e., the system, and not on its parts individually (e.g., Ottino, 2003). Here, a system can be viewed as a differentiated group of interacting elements jointly forming a complex and unified whole in the form of patterns that produce a behaviour to accomplish a specific purpose (Meadows, 2008).
- System properties like safety are not constant or resultant (Carayon et al., 2015). In fact, they are emergent, continuously arising from non-linear relationships among multiple parts of the system (e.g., Rasmussen, 1997; Leveson, 2004), which is defined by how the elements interact and fit together (Ackoff, 1971).

Therefore, a system state of safety cannot be achieved by studying the components taken separately, so optimisation of individual parts will not generally lead to system optimum. In fact, it may even worsen the system performance due to unintended and unexpected side-effects arising from complex and non-linear interactions, a phenomenon - “never change a running system” - which often has proven to be true over the long term (Leveson, 2002). Instead, a safe state is the emergent result of complex system flows from agents adapting their functioning to cope with changing conditions (Dekker, 2011; Dekker et al., 2011). For example, advanced driver assistance systems (ADASs) (e.g., ACC or LKA) are assumed to have a vast potential to improve road safety (Golias et al., 2002). However, individuals usually adjust their behaviour (e.g., increasing driving speeds, reducing the distance to a lead vehicle, paying less attention to the driving task) in response to changes in perceived risk due to the introduction of supporting technology which describes a phenomenon commonly known as risk compensation/risk homeostasis (e.g., Wilde, 1982; Fuller, 2005) or behavioural adaptation (OECD, 1990). That is also the reason why past success does not guarantee future success, as every system is unique and requires renewed evaluation of the whole system after the introduction of new elements, such as automation, because an evaluation just on the new sub-system level is not sufficient as other established sub-systems can change their behaviour. In addition, the obtuse introduction of automation makes a system more complex and intractable, leading to increased adaptations by people, leading to more unintended consequences and, in turn, even higher demand for automation, etc. – a circulus vitiosus (Hollnagel, 2016). Overall, the agents made local adjustments in their activities based on limited knowledge and resources to deal with the current situation, which can generate an unforeseen system evolution over time under normal or abnormal circumstances (Cooke & Rohleder, 2006). Thus, normal as well as adverse events emerge from the interdependencies among performance adjustments, which usually go well but rarely can fail, following normal functioning under normal circumstances (Pidgeon, 2010). For these reasons, the system properties have to be examined in a synthesised view, considering the multidimensional relationships between elements and the system in its entirety on a global level. Thus, it is commonly stated: "the whole is not the sum of its parts, it is the product of their interactions" (Awal street journal).

Various systems-based safety and risk management models have emerged over the past three decades (e.g., Hollnagel, 2004; Leveson, 2004; Rasmussen, 1997; Svedung & Rasmussen, 2002), which can be summarised based on a common set of systems thinking tenets (Grant et al., 2018) including 15 characteristics of complex work systems that are assumed to generate both safe and unsafe performance. Rasmussen's risk management framework is one popular systems theory-based model that started to be applied in road safety research (e.g., Newnam & Goode, 2015; Salmon et al., 2013; Scott-Parker et al., 2015; Young & Salmon, 2015). This framework describes a system as different hierarchical levels (e.g., government, regulators, company, management, staff, and work/activity), where each organisational level contains actors (individuals, organisations, or technology) contributing to production and safety management (see Figure 19).

Safety is seen as a control problem of hazardous processes where control is imposed on many levels, from the operational to the managerial (Leveson, 2002). For systems to function safely, a control-feedback loop is required where on the one hand, decisions at higher levels propagate top-down in the form of laws, rules, and instructions to be reflected in the decisions and actions at the lower levels, and on the other hand information at the lower levels is transferred bottom-up through observations, reports, and reviews to inform about the system status influencing the decisions and actions taken at the higher levels (Salmon et al., 2012). This vertical integration rather than horizontal orientation supports a system to control the processes it is designed to control (Rasmussen, 1997). The framework argues that decisions and actions at all system levels interact, shaping the system's performance. A key implication is that accidents are caused by multiple contributing factors, not just a single factor at the sharp end (e.g., an individual operator to be blamed) but multiple factors also involving the blunt end (e.g., managers or engineers). Accordingly, front-line workers in an accident usually represent symptoms rather than root causes, meaning latent failures at the blunt end are revealed by active failures at the sharp end. Thus, in most cases, it can be argued that the operator, immediate to the adverse event, as the final entity of different factors, could not prevent the accident. Therefore, Leveson (2004) claimed that the management's commitment to safety through a basic safety culture in the organisation is the crucial factor in the occurrence of accidents.

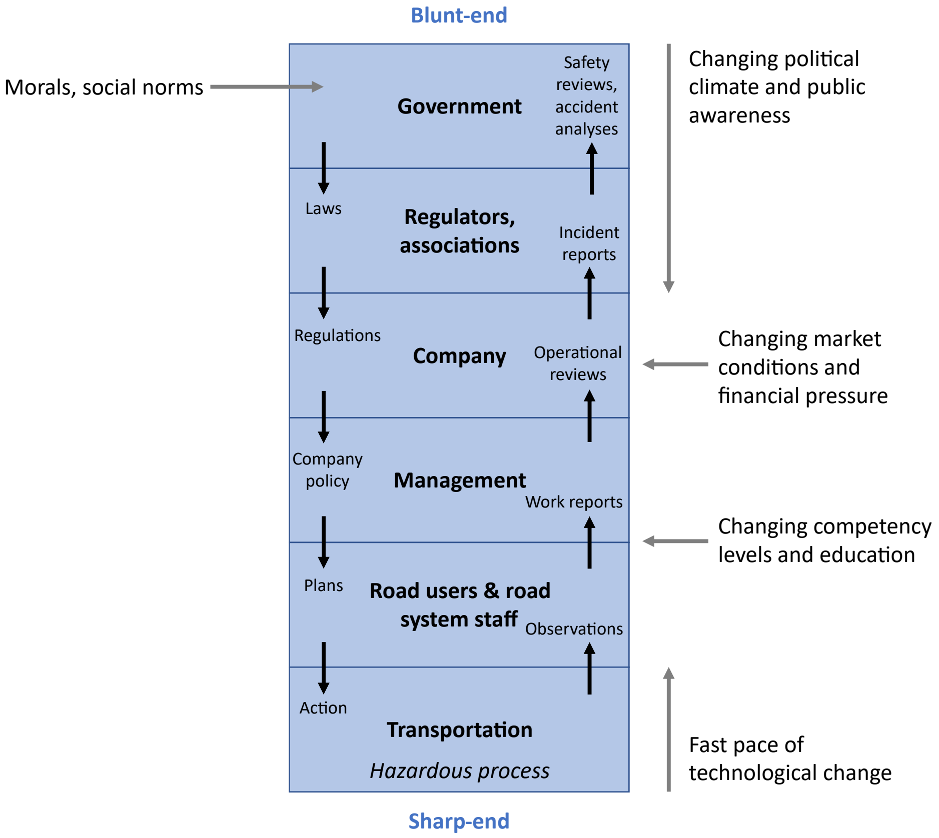

Blunt-end
Morals, social norms
Government
Safety reviews, accident analyses
Laws
Regulators, associations
Incident reports
Regulations
Company
Operational reviews
Company policy
Management
Work reports
Plans
Road users & road system staff
Observations
Action
Transportation
Hazardous process
Sharp-end
Changing political climate and public awareness
Changing market conditions and financial pressure
Changing competency levels and education
Fast pace of technological change

Figure 19. Rasmussen’s risk management framework adapted for road systems, adapted from Rasmussen (1997) and Read et al. (2017).

In addition, Rasmussen (1997) created the dynamic safety model (see Figure 20), arguing that work activities at all system levels are shaped by different objectives and constraints which actors must consider for successful work performance by adapting their behaviour which finally leads to a natural migration of activities over time toward the boundary of acceptable performance. The different targets and constraints (e.g., workload, cost-effectiveness, and safety represented as an "effort gradient", "efficiency gradient," or "safety counter gradient", respectively) define a "space of possibilities" in which individual actors navigate by resolving many degrees of freedom adjusting their performance. In addition, the objective and constraints are dynamic, so the space of possibilities continuously changes, making the required adjustments intractable.

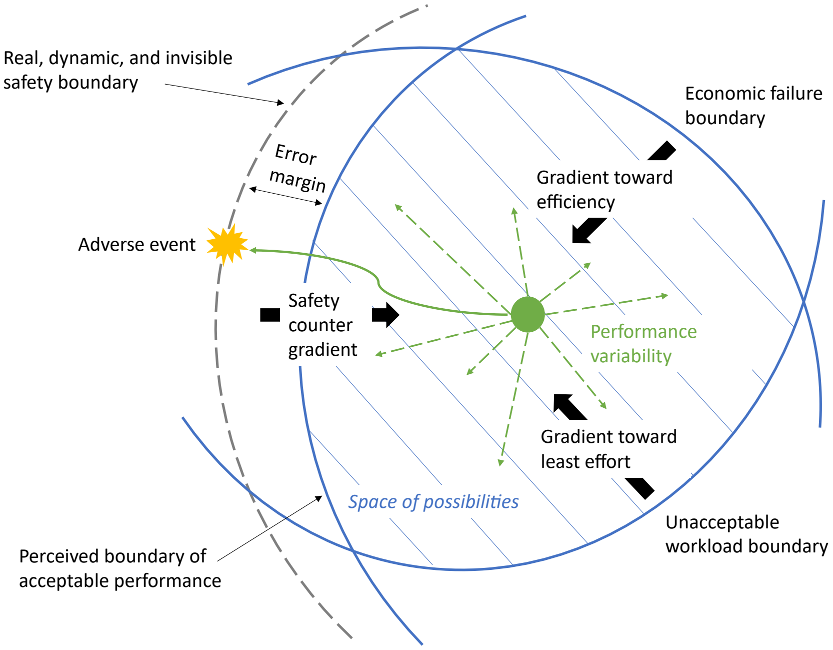

Real, dynamic, and invisible safety boundary
Error margin
Adverse event
Safety counter gradient
Economic failure boundary
Gradient toward efficiency
Performance variability
Gradient toward least effort
Space of possibilities
Unacceptable workload boundary
Perceived boundary of acceptable performance

Figure 20. Rasmussen's dynamic safety model, where the performance is migrating or drifting toward the boundary of acceptable performance, adapted from Rasmussen (1997) and Salmon et al. (2012).

In terms of road safety, this modelling approach is also related to the concept of "space of safe driving" by Gibson & Crooks (1938) or its revisited conceptual framework by Papakostopoulos et al. (2017). The performance adjustments are, among other external and internal performance shaping factors (Eurocontrol, 2009; Miller & Swain, 1987), conceptually guided by the principle of Efficiency-Thoroughness Trade-off (ETTO) (Hollnagel, 2009a). One explanation for these trade-offs is compensation, which means the need to absorb the effects of the everyday performance variabilities made by the system's remaining actors (Hollnagel, 2012a p.31). The local variations induced by situational conditions show a great performance variability calling in mind "brownian movements of the molecules of a gas" (Rasmussen, 1997). Hollnagel (2004) emphasises that these variabilities are quite normal, especially locally, rather than inherently bad or abnormal. In fact, they are necessary for a system to match current demands and resources, which are dynamic and not entirely predictable, to fulfill its purpose. Unfortunately, under certain conditions, these variabilities can rarely evolve in a manner that leads to a crossed functional safety boundary, which is irreversible, and an error or accident might occur. The real, functional safety boundary is dynamic and invisible, resulting in a flexible error margin that makes it challenging to operate at the limits of safe states or acceptable performance.

Like Rasmussen's work, Dekker (2011) proposed the drift into failure model, describing how multiple decisions and actions over time in different dimensions with limited knowledge of system-wide effects gradually shift the complex system performance, often unnoticed, towards harmed safety or adverse events. According to Rasmussen (1997), accidents are typically "waiting of release". This means that a stage of accidental course is being set through time by routine activities in the daily work context where a normal variation in behaviour then might release the accident. That is the reason why the traditional approach explaining accidents as deterministic cause-effect relationships in terms of events or errors is not expedient to design improved systems because avoiding supposed root-causes, such as individual behaviour deviations, by additional safety measures would likely release an accident by another cause at another point in time and space (Rasmussen, 1997). Similarly, Perrow (1984) argues that accidents are inevitable and unavoidable in highly complex systems, calling it the "normal accident theory". In line with that, Rasmussen (1997) claimed that for a general understanding of system behaviour, we do not have to focus on structural decomposition but rather on functional abstraction and mechanisms shaping the behaviour in the system embedded in the actual, dynamic work context within the degrees of freedom.

However, safety management usually takes a retrospective approach focusing only on adverse events like accidents. In particular, accident investigations suffer a "hindsight bias" by knowing all consequences and thus more than the actual actors involved in the accident, which leads to looking at certain things and limiting an open mind during the analysis (Lundberg et al., 2009). This reasoning is associated with the so-called What-You-Look-For-Is-What-You-Find (WYLFIWYF) principle (Hollnagel, 2008), leading to the What-You-Find-Is-What-You-Fix (WYFIWYF) principle (Lundberg et al., 2009). Hence, causes are not found but rather constructed and selected (Dekker et al., 2011) following the credo to control behaviour by tackling deviations from prescribed instructions or plans (Rasmussen, 1997), also called work-as-imagined (WAI) (cf. Dekker, 2006; Ombredane & Faverge, 1955) which is inherently flawed. Instead, the focus should be to understand why a specific action or decision taken at a particular space-time continuum made sense to the person who has taken this action or decision (Dekker et al., 2011) acknowledging work-as-done (WAD) (cf. Dekker, 2006; Ombredane & Faverge, 1955) affected by inevitable performance adjustments due to the complexity and then helping people to develop skills for coping with processes within or at boundaries (Rasmussen, 1997). The most closest "truth" about an accident may only become visible by considering multiple narratives from different perspectives rather than a single one (Dekker et al., 2011).

This relates to the concept of resilience, which has been defined as "the intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances to continue working in the face of continuous stresses or major mishaps" (Hollnagel et al., 2006; Nemeth et al., 2008). Hence, RE, since the first Resilience Engineering Association (REA) symposium held in 2004 (Dekker, 2006), constitutes a paradigm shift in safety management by focusing on "guided adaptability" (Cook et al., 1998; Hollnagel, 2014) in contrast to "centralised control" (Provan et al., 2017). The concept of adaptation is based on four cornerstones: responding (knowing what to do), monitoring (knowing what to look for), anticipating (knowing what to expect), and learning (knowing what has happened) (Hollnagel, 2009b). Furthermore, RE provides means to manage risks proactively (Woods, 2003) and thus enhance the system's capability to cope with complexity due to balancing productivity with safety in everyday work in response to normal and abnormal operating conditions (Dekker et al., 2008; Hollnagel, 2006a).

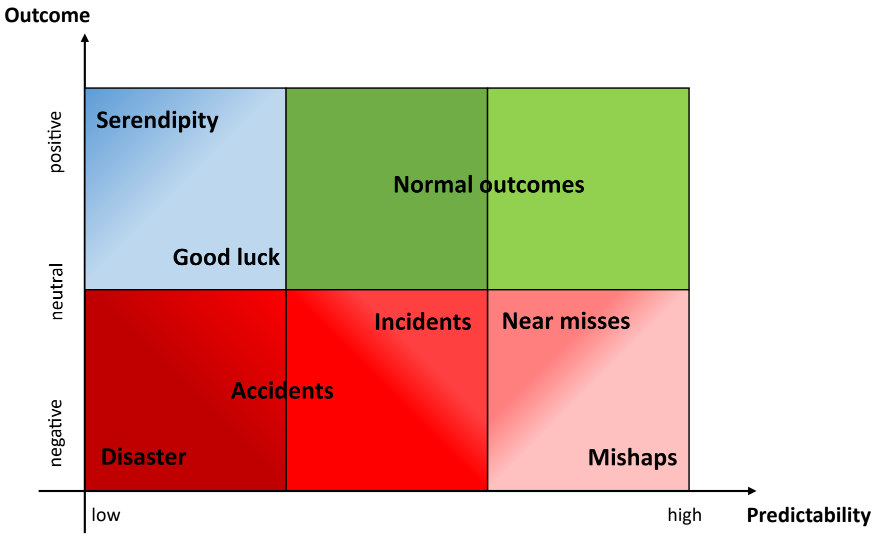


Figure 21. The set of possible outcomes adapted from Eurocontrol (2009).

RE is based on the purpose and assumptions of safety-II, as mentioned in Section 2.5. Hence, safety becomes a positive and productive meaning covering the whole set of outcomes distinguishing between the two dimensions of predictability and outcome ranging from low to high and negative to positive, respectively (Eurocontrol, 2009) (see Figure 21). Moreover, the database, which we can learn from, is much broader when assuming a normal distribution of outcomes (Shorrock, 2022) (see Figure 22**Figure *21***). However, the distribution may vary (e.g., skew or kurtosis) due to different conditions and changes over time, but only with a marginal decrease of data related to everyday performance (Shorrock, 2022). According to Hollnagel (2014), it makes more sense to analyse small but frequent events (everyday performance) instead of large but rare events (accidents) because the former are easier to understand and manage and also have more impact on the safety of the overall system. It must be emphasised that bad and exceptional performance are not opposites but rather closely related (Shorrock, 2022). In terms of the dynamic safety model, this means that, in case of exceptional performance, the operating state is close to the boundary of acceptable performance but still within the safe space of operation, i.e., normal everyday performance. When crossing the boundary of acceptable performance, the outcome becomes bad performance showing up as a near miss or incident, ultimately resulting in accidents when crossing the functional safety boundary.

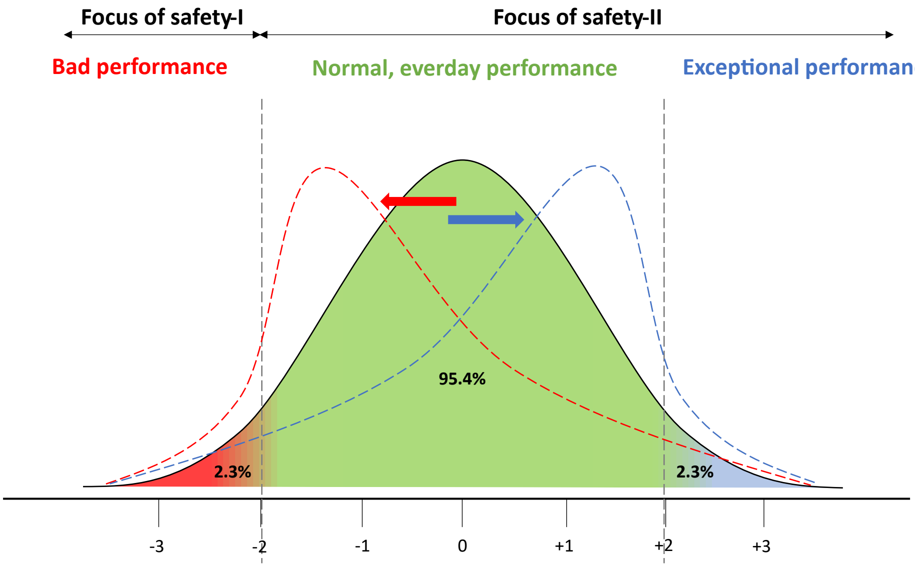


Figure 22. The theoretical normal distribution of outcomes/performances adapted from Shorrock (2022).

As we can see, RE aims to comprehend why a system mostly works in order to understand how it can rarely fail. In order to achieve this, the focus is to fully understand the adaptive capacity of the qualities of mechanisms enhancing the system's resilience (Patriarca, 2017). Ultimately, RE may play a pivotal role in an era of increasingly more complex systems (Patriarca, 2017).

4 Research Objective

> "Successful problem solving requires finding the right solution to the right problem. We fail more often because we solve the wrong problem than because we get the wrong solution to the right problem."
>
> – Russel Ackoff –

4.1 Problem definition and objective

Automated driving promises great possibilities for traffic safety advancement. However, as we have seen in the previous sections, the safety of automated vehicles is the subject of controversial discussion, and no safety proof of HAD has yet been provided due to challenges in the approval process resulting in the approval trap (Winner, 2016). Therefore, new test methods and risk evaluation methods must be developed. However, as previously indicated, the commonly used methods heavily rely on reductionist and mechanistic reasoning, which are limited useful and increasingly inadequate to assess the safety of HAD in road traffic. Rather a systemic and complexity-oriented, holistic approach must also be applied, which is almost neglected. In particular, a more differentiated view is lacking that tries to understand the inherent adaptation processes in complex STSs, i.e., the road traffic system. This view also requires improving the safety of the entire road system by focusing on the efficient interaction between humans, machines, and other road users (Bengler et al., 2017). Overall, the processes in the road system leading to accident development and accident avoidance have to be differentiated, including the interdependencies between each element in the system. According to Rasmussen (1997) and Patriarca (2017), this belongs to the comprehension of the mechanisms characterised by adaptive capacity in the system which shapes the behaviour to create resilience. However, as the word "mechanism" could be misleading as implying a newtonian, mechanistic reasoning which represents a paradox to the safety argumentation and goal of this thesis, it should be replaced by the word "pattern".

Patterns represent how sharp-end, as well as blunt-end agents, adapt their behaviour to cope with the complexity of the work (Eurocontrol, 2021a; Hollnagel & Woods, 1983), i.e., how activities or functions in a system are carried out (Shorrock, 2016). To express it more formally, a pattern captures a set of relationships between elements where the pattern emerges from these interactions representing emergent properties not present in the elements (Eurocontrol, 2021b). If we recognise and understand those patterns, we can use them to understand what happens and anticipate what may happen in the future (Hollnagel, 2016). In general, adaptations make a system function but hide its weaknesses, so they are often overseen (Eurocontrol, 2021b). This leads to the distinction between strong and weak signals *(*Eurocontrol, 2021a*).* Strong signals are distinctive and well-defined events such as incidents or accidents which are difficult to miss as they are clearly visible as a signal located significantly above a detection threshold. In contrast, weak signals seem to be disconnected pieces of information comprising small, subliminal events that usually keep unreported and recurrent performance patterns over time, e.g., habits, routines, and trade-offs, usually resulting in expected outcomes but seldomly unwanted outcomes. This is similar to the dynamic non-events that are the foundation of reliable performance (Weick, 2011). A key insight by the "pattern-centered inquiry" from Alexander et al. (1977) is that a pattern is rather general but expressed in many different situations and settings, meaning that if the weaknesses in a system, revealed by pattern identification, are correctly addressed through system design measures, then a huge positive impact on the system performance can be expected. In particular, RE is based on the pattern approach by identifying how adaptations in a system work and what drives these adaptation processes (Patterson et al., 2007) in order to develop empirical patterns of adaptive behaviour (see, e.g., Woods, 2019; Woods & Branlat, 2017).

In formal terms, this approach refers to the data, information, knowledge, and wisdom (DIKW)-pyramid for knowledge management (Ackoff, 1989) (see Figure 23). This framework represents a hierarchy of knowledge gain starting from data collection and then enriching data with context, meaning, and insight so that data is converted upwards into information, knowledge, and wisdom. Each level is a step forward in understanding and connectedness, which supports decision-makers to make better decisions, e.g., in safety management to enhance safety. At the bottom level of knowledge management, the analysis takes place by decomposing systems into separated parts characterised through hindsight. In contrast, at the top level, a synthesised view is taken by holistic studying of joint systems, which is then used proactively for the future. Wisdom can only be reached by increasing pattern identification. This means that fragmented knowledge, evolved by conversions of tacit and explicit knowledge in the form of socialisation, externalisation, combination, and internalisation (SECI-model) (Nonaka et al., 2000), is elicited, jointly studied and combined into a multifaceted perspective to capture nuances of work that guides system improvements and redesigns (Eurocontrol, 2022).

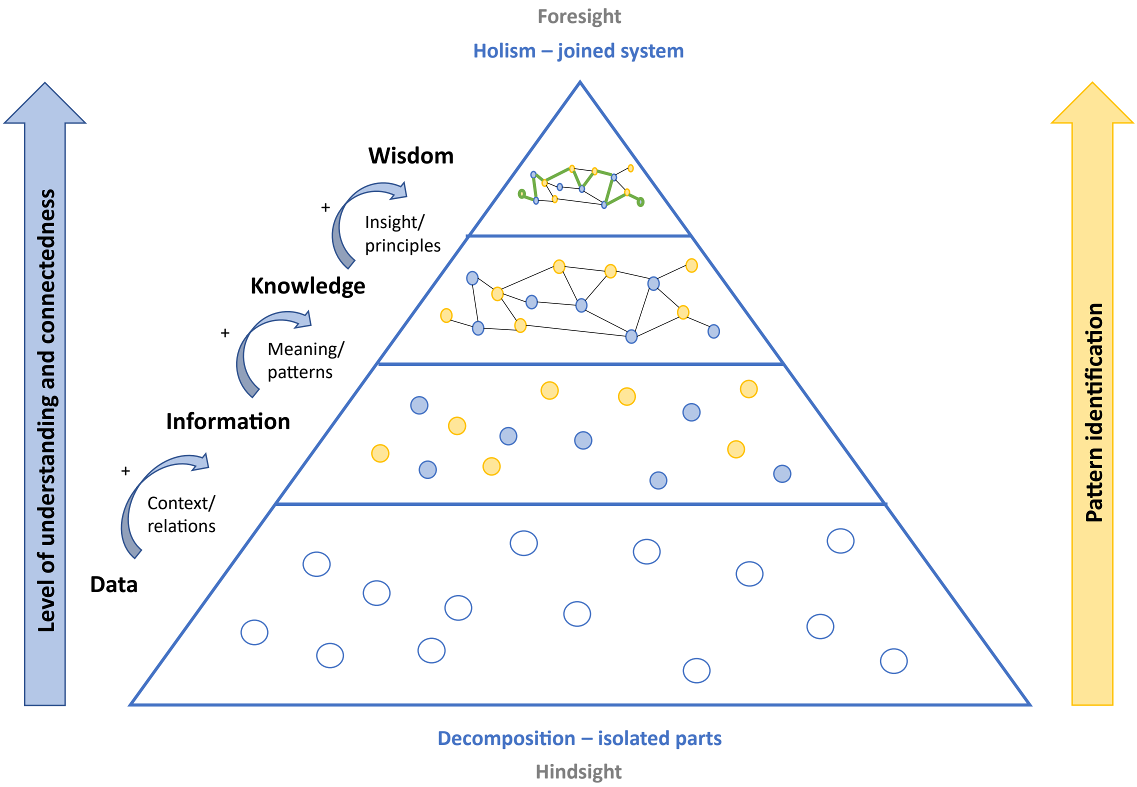


Figure 23. The DIKW-pyramid for knowledge management adapted from Ackoff *(1989), Cannas et al. (2019), Eurocontrol (2022), and Flood et al. (2016).*

Ultimately, the word "pattern" can be used more broadly, including mechanisms and emergentisms, which Walker et al. (2009) argued are contributing both to an STS's performance. A mechanism can be defined as a process or system of elements interacting in a fixed, predictable way following linear cause-effect relationships resulting in a resultant behaviour. In contrast, emergentisms represent the opposite and are defined in this thesis as a pattern or system of elements interacting in a complex and dynamic way following non-linear and elusive cause-effect relationships resulting in emergent behaviour. In particular, the emergentisms in road traffic are in focus as they represent the added value that is currently lacking compared to mechanisms analysed by the traditional approaches, such as Heinrich's domino model, Ishikawa diagrams, Reason's swiss cheese model, and Leveson's STAMP.

4.2 Research questions, assumptions, and further outline

Therefore, the final aim of this thesis is to create a differentiated understanding of the patterns in road traffic leading to accident development and accident avoidance (see Figure 24). This goal will be exemplified by one scenario to demonstrate the methodological approach on a small-scale, which has to be applied on a large-scale in the future. Based on this, the contribution of both the driver and the automation to the system performance can be assessed to derive design recommendations for the system and validation process within one scenario, especially in terms of safety-II. Before these patterns can be identified, first, a suitable method, i.e., FRAM, has to be identified and methodologically evaluated, which can reveal patterns assessing road safety related to human and automated driving, and second, reasonable test scenarios have to be defined for the safety assessment of automated driving in which FRAM should be applied. Then, in a third step, FRAM is used in a specific scenario by creating a model to understand the patterns of accident development and accident avoidance to give system design recommendations and essential insights for the validation process. In a fourth and final step, the model and method are evaluated in terms of validity to assess the credibility of the results and the applicability, respectively.

These four research steps are described in Sections 5-9. The research questions are based on the following assumptions: safety is assessed in terms of safety-II, and the assessment compares manual drivers (LoDA 0) with HAD (LoDA 4/5) in mixed traffic. In this thesis, mixed traffic means a mix of manual drivers and HAD but no other levels of automation and no complete penetration by HAD.

Ultimately, the results of this research process are integrated into the overall discussion of this thesis (Section 10) to:

- derive system design and validation recommendations, including possible technical implementations as well as criticism of the current LoDA,
- arguing unique features of FRAM but also methodological limitations to demonstrate its value compared to other methods in the product development cycle of automated vehicles,
- and illustrating a potential application in the field or industry.

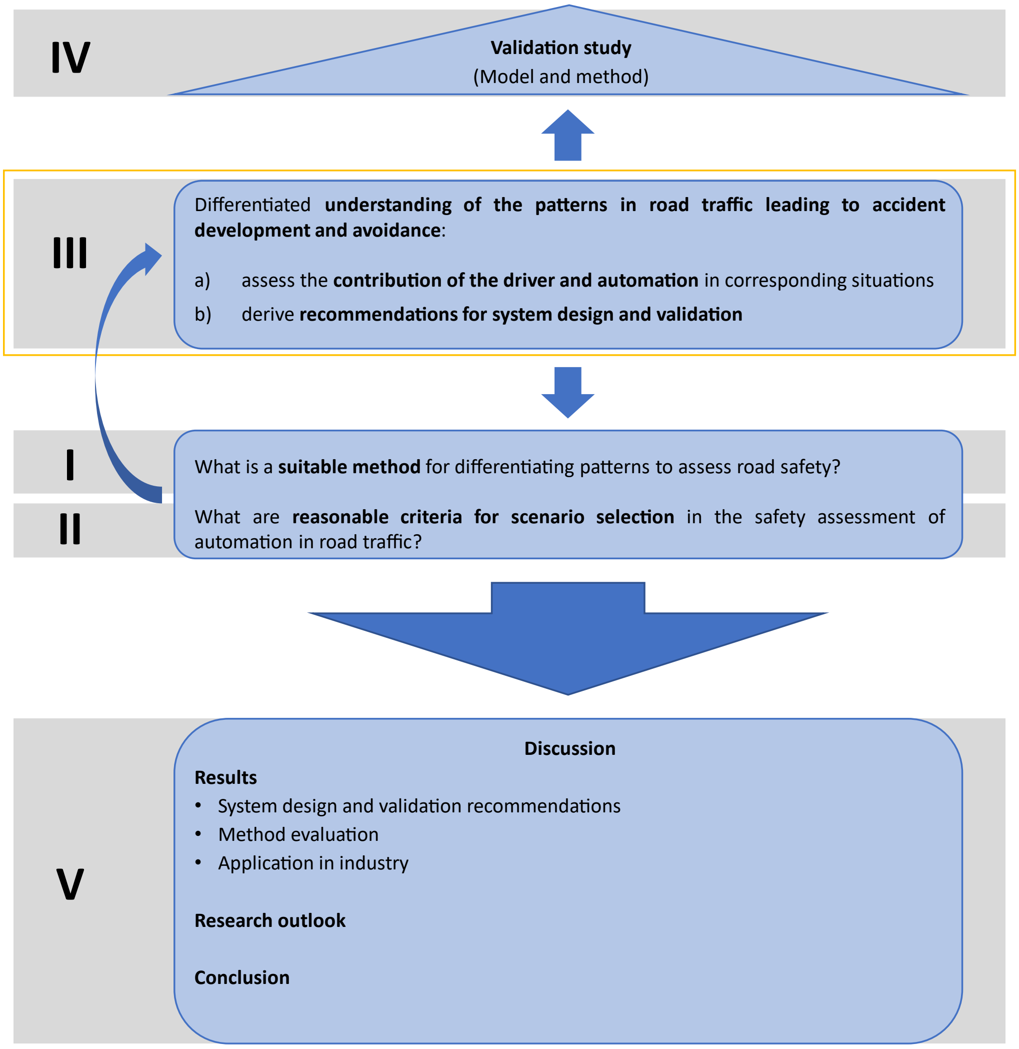

IV
Validation study
(Model and method)
III
Differentiated **understanding of the patterns in road traffic leading to accident development and avoidance**:
a) assess the **contribution of the driver and automation** in corresponding situations
b) derive **recommendations for system design and validation**
I
II
What is a **suitable method** for differentiating patterns to assess road safety?
What are **reasonable criteria for scenario selection** in the safety assessment of automation in road traffic?
V
Discussion
Results
System design and validation recommendations
Method evaluation
Application in industry
Research outlook
Conclusion

Figure 24. Overview of the research outline.

5 Article 1: "Safety of automated driving: the need for a systems approach and application of the Functional Resonance Analysis Method"

Grabbe, N., Kellnberger, A., Aydin, B., & Bengler, K. (2020). Safety of automated driving: the need for a systems approach and application of the Functional Resonance Analysis Method. *Safety science, 126*, 104665. https://doi.org/10.1016/j.ssci.2020.104665

Author Contributions: Conceptualisation, N.G.; methodology, N.G., A.K., and B.A.; software, N.G., B.A.; validation, N.G., A.K., and B.A.; formal analysis, N.G., A.K., and B.A.; investigation, N.G., A.K., and B.A.; resources, N.G.; data curation, N.G., A.K., and B.A.; writing—original draft preparation, N.G.; writing—review and editing, N.G.; visualisation, N.G.; project administration, N.G.; supervision, K.B. All authors have read and agreed to the published version of the manuscript.

> "The way the parts fit together determines the performance of the system and not on how they perform taken separately."
> – Russel Ackoff –

Summary

This article illustrates the challenges of assessing the safety of automated vehicles, represented as the approval trap (Winner, 2016), and provides a brief overview of idiosyncrasies in the safety argumentation of automated driving. Obviously, new test methods must be developed focusing on the differentiated understanding of the mechanisms and emergentisms of road traffic leading to accident development and avoidance. In particular, the method should facilitate identifying the driver's and automation's contributions to road safety. Therefore, the authors argue in favour of FRAM as a risk assessment method in the early stage of the development process of highly-automated vehicles (see Figure 10), primarily to derive system design recommendations and secondly to provide essential insights into reducing the validation work.

It begins with a systematic derivation of the benefits and suitability of FRAM. Here, from a theoretical standpoint, systemic methods and models are best suited to assess road safety against the background from introduced HAD because road traffic is a complex STS, as also comprehensively discussed in Sections 2.5 and 3. Thus, the most common systemic methods, Accimap (Svedung & Rasmussen, 2002), STAMP (Leveson, 2004), and FRAM (Hollnagel, 2012a), are methodologically compared against several aspects, e.g., systems-based characteristics.

FRAM is then applied to an overtaking manoeuvre on a rural road in a road traffic illustrative case study to evaluate its suitability and applicability in more detail, following the typical four steps of FRAM, i.e., functions identification, functions' performance variability manifestation, aggregation of variability, and management of variability. It should be noted that functions in FRAM describe activities or processes whose produced outputs are coupled to achieve a system goal. It is explicitly described for each function in the form of "production rules" (internal processes) of how these outputs are generated. This enables to create a white-box model to understand the inner workings of a system rather than a black-box model focusing on the outcome of the input-output relation.

The situational analysis of the behavioural requirements of driving tasks (SAFE) (Fastenmeier & Gstalter, 2007) and the functional decomposition of the road system (Kuzminski et al., 1995) were applied as the fundamental basis to identify and define the functions for the driving tasks (WAI) or activities (WAD) in FRAM. Then, the functions and their couplings were integrated iteratively into an instantiated model using the software *Functional Model Visualiser* (FMV) (Hill & Hollnagel, 2016). It should be noted that instantiation means the transfer of the relational organisation between functions from a potential to an actual sequence (upstream-downstream coupling) and type of relationship (e.g., output–input coupling), i.e., temporal and causal relations that might occur in a different way each time (Hollnagel, 2012a) which is why instantiations change depending to the conditions. In this thesis, for the sake of simplicity, we use the term "FRAM model" as an "instantiation of an FRAM model". In the second step, the manifestation of variability was determined subjectively. Afterwards, the aggregation of the variability was implemented using an enhanced semi-quantitative approach by adapting Patriarca's framework based on Monte-Carlo simulation (2017b) due to high complexity which is difficult to handle when applying FRAM in its traditional, qualitative way. Finally, in the fourth step, the functional resonance between the driver and automation is compared by analysing how the variability may propagate through the system, creating functional resonance using exemplary calculations and critical paths.

Ultimately, a discussion of the first application of FRAM to the road system follows, presenting FRAM's strengths and limitations (see Table 1). FRAM enables identifying critical functions and their consequences for the entire system and visualising mechanisms and emergentisms illustrating interaction patterns in road traffic. The article concludes that FRAM supports decision-makers in enhancing safety enriched by identifying non-linear and complex risks rather than the linear cause–effect-related risks that are frequently the sole focus of safety and risk assessments at present. Finally, the conclusions consider FRAM as a missing piece in the puzzle for a proactive risk assessment of automated driving and its system design, illustrating the need for further research due to limitations.

Table 1. Overview of methodological strengths and limitations of FRAM.

Strengths	Limitations
• Flexible and agnostic method without limited model assumptions (method-sine-model), WAD • Oppenness, "toy-model" • Guidance material • Software (standardisation) • Graphical representation • Integration of qualitative as well as quantitative data • Facilitation to comprehend the complexity	• Elaborate, intensive training and much previous knowledge • Qualitative representation quickly overwhelming but compensated through semi-quantitative approaches • Unclear strategies to identify functions and their variability • (Subjective) modeling, calibration instead of validation

6 Article 2: “Safety Enhancement by Automated Driving: What are the Relevant Scenarios?”

Grabbe, N., Höcher, M., Thanos, A., & Bengler, K. (2020). Safety Enhancement by Automated Driving: What are the Relevant Scenarios? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 64, No. 1, pp. 1686-1690). Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/1071181320641409

Author Contributions: Conceptualisation, N.G; methodology, N.G., M.H, and A.T.; validation, N.G., M.H, and A.T.; formal analysis, N.G., M.H, and A.T.; investigation, N.G., M.H, and A.T.; resources, N.G.; data curation, N.G., M.H, and A.T.; writing—original draft preparation, N.G.; writing—review and editing, N.G.; visualisation, N.G.; project administration, N.G.; supervision, K.B. All authors have read and agreed to the published version of the manuscript.

> “Sometimes a change of perspective is all it takes to see the light.”
> – Dan Brown –

Summary

This article addresses the scenario-based approach to solve the approval trap (Winner, 2016) by reducing the test scope toward relevant or crucial scenarios based on reasonable criteria for scenario selection. Unfortunately, the current approach still results in a huge number of test cases (Amersbach & Winner, 2019). The challenge is to find a set of representative but still efficient scenarios suitable for scenario-based approval. Riedmaier et al. (2020) provide a comprehensive overview of several approaches to identify and select these scenarios. Two main selection processes that follow the ground scientific reasoning principles of induction and deduction can be differentiated: testing-based to cover the parameter space with finite test-cases or falsification-based by focusing on challenging edge-cases to find counterexamples. However, even if all test-cases are successfully passed, and no counterexamples can be found, safety still cannot be guaranteed entirely because both procedures are microscopic assessments by using key performance indicators (KPIs), such as TTC, applied to specific cases, and conclusions cannot be drawn on a macroscopic level for the entire system. One possible way out is to show the current, incorrect path in the argumentation and strategy of vehicle automation, as described in Sections 2.5 and 3, and rather focus on the systemic patterns of road traffic safety. Therefore, this paper argues the case for defining relevant, abstract scenarios in mixed traffic and analysing them systemically in terms of safety-II rather than following reductionism in terms of safety-I to reduce the test cases ultimately. Thus, a microscopic and macroscopic assessment are jointly combined into one approach.

The relevant scenarios are knowledge-driven and based on the drivers' and automation's strengths and weaknesses in the driving task. Two different types of scenarios are distinguished following the suggestions by Bengler et al. (2017). Based on accident statistics, type-I describes scenarios that offer great potential for significant safety improvement through automation because humans have proven highly likely to contribute to accident occurrence, i.e., accident black spots and risk groups of drivers. Type-II comprises scenarios that represent either the unique strengths of the driver in uncritical and accident-free situations or supposed challenges for automation. This is based on a succinct synthesis of literature as well as expert interviews.

Finally, abstract, basic rather than explicit, concrete scenarios as criteria for exclusion, like a falsification-based approach, are being proposed to systemically assess the contribution of the driver and automation to road safety. According to the abstraction levels of scenarios (Menzel et al., 2018), the scenarios represent rather functional than concrete scenarios. The "clever trick" is not to vary specific parameters systematically and derive concrete scenarios, and then test each specific scenario by analysing KPIs concerning events but instead gather everyday performance data of several driving activity-related functions in varying conditions in real traffic within one abstract scenario resulting in a distribution function for each function which then has to be studied systemically and holistically by FRAM analysing resilience indicators or metrics of the entire system performance. Thus, it is less important to pay attention to critical events such as errors or accidents than to focus on the variability in the performance of the individual driving activities in uncritical and normal driving and predict their potential propagation effects in the system.

This also means that significantly fewer test kilometres must be covered since data can be gathered immediately (see Figure 25). For safety-I, we have a huge test distance but only one data point or a bimodal event distinguishing an accident and no accident. Instead, for safety-II, we have a significantly reduced test distance but a massive increase in data amount and quality. Thus, there is no need to wait for an accident to occur or something bad to happen because anything can be measured at any time. Rather, we have to understand what actually happens in situations where nothing out of the ordinary seems to take place, such as dynamic non-events or weak signals, and to compare this between the driver and the automated vehicle. Hence, it is sufficient to develop a description of the daily activity and its expected variability, which means one generic case instead of many specific ones. Therefore, it makes more sense to analyse small but frequent events (everyday performance) instead of large but rare events (accidents) because the former are easier to understand and manage and also have more impact on the safety of the overall system (cf. Hollnagel, 2014). Clearly, this will result in a large amount of data that must be analysed automatically. Fortunately, this problem should be solvable compared to the extrapolated lots of test kilometres by Wachenfeld & Winner (2016), based on track distance that currently focuses only on accidents or incidents.

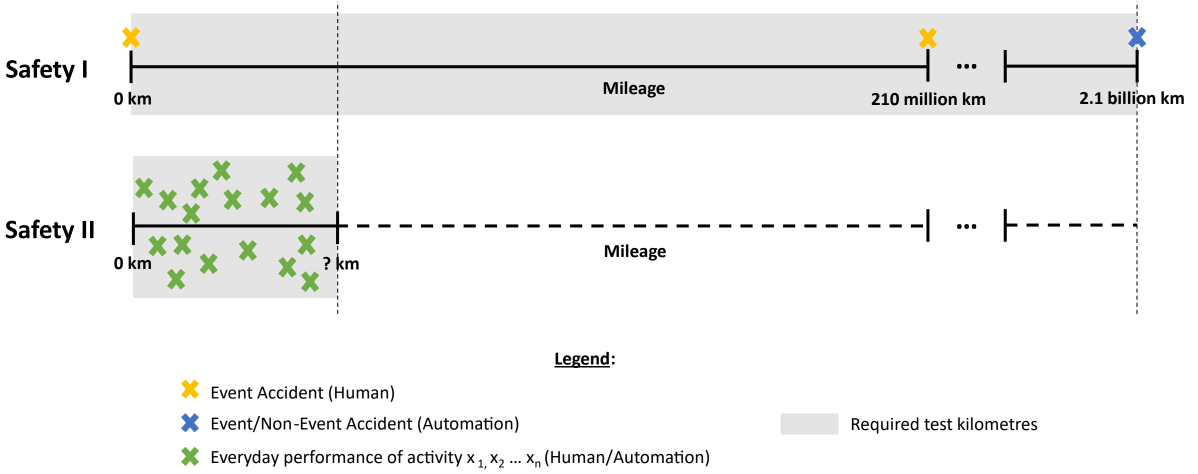


Figure 25. Comparison between the safety-I and safety-II view on road safety testing based on the two parameters test distance and data amount. The test distance in the safety-I approach is based on the calculations by Wachenfeld & Winner (2016) for driving on average.

Ultimately, it is concluded that the derived scenarios do not claim to be complete. However, with the presented relevant scenarios as decisive factors, a solid foundation to systemically analyse these scenarios is set in order to build an understanding of the system's interrelationships and its actual patterns that are needed as key insights to support the design of safe automated vehicles proactively and to reduce the validation work.

7 Article 3: "Functional Resonance Analysis in an Overtaking Situation in Road Traffic: Comparing the Performance Variability Mechanisms between Human and Automation"

Grabbe, N., Gales, A., Höcher, M., & Bengler, K. (2022). Functional Resonance Analysis in an Overtaking Situation in Road Traffic: Comparing the Performance Variability Mechanisms between Human and Automation. *Safety, 8*(1), 3. https://doi.org/10.3390/safety8010003

Author Contributions: Conceptualisation, N.G., A.G., and M.H.; methodology, N.G., A.G, and M.H.; software, N.G., A.G.; validation, N.G., A.G., and M.H.; formal analysis, N.G., A.G., and M.H.; investigation, N.G., A.G., and M.H.; resources, N.G.; data curation, N.G., A.G., and M.H.; writing—original draft preparation, N.G.; writing—review and editing, N.G.; visualisation, N.G.; project administration, N.G.; supervision, K.B. All authors have read and agreed to the published version of the manuscript.

„ We should work on our process, not the outcome of our processes."
– W. Edwards Deming –

Summary

This article continues the research of article one in Section 5 (Grabbe et al., 2020b) to reduce the research gap in the safety assessment of automated vehicles regarding the safety-II perspective. The aim is to identify road traffic patterns contributing to safety in an overtaking scenario that represents a huge potential to increase safety through automation in a complex setting (cf. Grabbe et al., 2020a) using FRAM. The contributions between the driver and automation are compared to derive system design recommendations. Finally, this demonstrates how FRAM can be used for a systemic function allocation for the driving task between humans and automation.

Thus, an in-depth, instantiated FRAM model was developed for both agents based on document knowledge elicitation and observations and interviews in a driving simulator, which was validated by a focus group with peers. Further, the performance variabilities were identified by structured interviews with drivers, automation experts,

and observations in the driving simulator. Then, the aggregation and propagation of variability were analysed, focusing on the interaction and complexity in the system by an extended semi-quantitative approach combined with a Space-Time/Agency framework and enhanced analysis metrics. Finally, design recommendations for managing performance variability were proposed through a well-reasoned function allocation to enhance system safety. To achieve this, the performance variability of the entire system is analysed by comparing the contributions between driver and automation to road safety based on patterns on both an abstract global level and a fine-grain level regarding the individual functions.

The design recommendations for function allocation between driver and automation can be seen as a JCS (Hollnagel & Woods, 2005) that regards human and machine as equal partners collaborating in the sense of a human-machine coagency which is expressed in terms of function-centeredness (Hollnagel, 2006b) where system functions needed to accomplish the overtaking manoeuvre are distributed between the driver and/or the automation in consideration of the interactions and dynamics in a space-time continuum within and between agents in the system reflected by system resonance and the functional variabilities. The outcomes show that the current automation strategy should focus on adaptive automation based on a human-automation collaboration rather than full automation. In conclusion, RE, in particular, FRAM, can be applied to the road traffic system to proactively and holistically design automated driving functions as a joint driver-vehicle system that supports decision-makers in enhancing safety enriched by identifying non-linear and complex risks.

8 Article 4: "Assessing the reliability and validity of an FRAM model: the case of driving in an overtaking scenario"

Grabbe, N., Arifagic, A., & Bengler, K. (2022). Assessing the reliability and validity of an FRAM model: the case of driving in an overtaking scenario. *Cognition, Technology & Work, 24(3),* 483-508. https://doi.org/10.1007/s10111-022-00701-7

Author Contributions: Conceptualisation, N.G., A.A.; methodology, N.G., A.A.; software, N.G., A.A.; validation, N.G., A.A.; formal analysis, N.G., A.A.; investigation, N.G., A.A.; resources, N.G.; data curation, N.G., A.A.; writing—original draft preparation, N.G.; writing—review and editing, N.G.; visualisation, project administration, N.G.; N.G.; supervision, K.B. All authors have read and agreed to the published version of the manuscript.

„All models are wrong, but some are useful."
– George E.P. Box –

Summary

This article contributes to the current lack of any formal testing of the reliability and validity of FRAM, which applies to Human Factors and Ergonomics (HFE) research as a whole, where validation is both a particularly challenging issue and an ongoing concern. The goal is to define a more formal approach to achieving and demonstrating the reliability and validity of an FRAM model, as well as to apply this formal approach partly to the instantiated FRAM model created in article three in Section 7 (Grabbe et al., 2022b) to prove its validity. At the same time, it hopes to evaluate the general applicability of this approach to improve the performance and value of the FRAM method.

Thus, a formal approach or framework was derived by transferring the general understanding and definitions of reliability and validity and concrete methods and techniques to the concept of FRAM. Consequently, predictive validity, the highest validation maxim, was assessed for the specific FRAM model in a driving simulator study using a mixture of outcome-based and function-based validation, including the signal detection theory combined with a what-if-analysis. In particular, two functions of

the model were manipulated in varying environments and human factors conditions to see if the predicted changes and non-changes in performance variability of other affected functions can be observed in reality.

It is concluded that the FRAM model's predictive validity is limited, particularly in its specificity, indicating deficiencies in the credibility of the examined FRAM model. Moreover, the generalisation with changing system conditions is impossible without some adaptations of the model. However, this is not surprising as an FRAM model can

only be validated for specific instantiations, and if the conditions change, the instantiation will change. The model must then be adapted, and no generalisation will be possible. Overall, the developed framework provides a good foundation to evaluate the reliability and validity of an FRAM model, especially helping analysts compare FRAM's cost-effectiveness with other HFE methods. Ultimately, the applicability of the approach is diminished because of several methodological limitations. Therefore, the reliability and validity framework can be utilised to calibrate rather than validate an FRAM model.

9 Articles 3 and 4 revisited: an enhanced pattern identification combined with a function-based validation approach

"Wisdom consists of the anticipation of consequences."
– Norman Cousins –

9.1 Introduction

As we have seen previously in Section 8 (Grabbe et al., 2022a), the outcome-based validation to evaluate the predictive validity of the examined FRAM model is limited. Therefore, a pure function-based validation approach is applied by a sensitivity analysis with deliberate and controlled variations in the model as a falsification approach. Specifically, the response mode of the model is checked for plausibility to get an enhanced comprehension of the model's credibility. These variations represent a changing automation of different agents and functions in the FRAM model, extending the overtaking scenario used in Grabbe et al. (2022b). In principle, different instantiations of the FRAM model are compared in the sense of scenario-based envisioned systems. Moreover, apart from an improved cost-effectiveness trade-off, the function-based validation has the benefit of evaluating the predictive validity of the entire FRAM model rather than one part of the model when merely using the outcome-based validation approach. However, it should be emphasised that FRAM is not comparable with the better-performing HFE methods which typically achieve validity statistics above 0.8 (cf. Stanton et al., 2022). The reason is that FRAM, compared to other HFE methods, depending on the chosen granularity level of modelling, predominantly covers the prediction of system behaviour more extensively and for a more abstract and holistic system view. Instead, typical HFE methods are pretty special, applied to narrowed cases offering limited insights on a decompositional level. Thus, FRAM has the inherent property to have a higher chance or risk of false predictions but is compensated through a high potential for insights concerning the holistic system level that are correctly predicted.

The theoretical backdrop is as follows. It is frequently assumed that the safety of the whole traffic system is improved by automated vehicles if the HAV and other traffic participants are driving compliant. However, this is an unrealistic assumption, especially in situations with much interaction, as the road system is an open system, including deviations from the norm with required adaptations. Thus, the theory above would only be valid for a closed system consisting of merely HAVs. It is therefore hypothesised that introducing HAVs into mixed traffic in an overtaking scenario will destabilise the overall system functioning. Thus, the system is more stable with only manual drivers or a joint collaboration between humans and automation. The assumptions are as follows:

- the system remains the same, independent from automation levels, which is why the functional structure of the FRAM model remains the same; hence, only performance variability values will change
- no V2X communication is implemented, and the capabilities of HAD represent the current state-of-the-art
- mixed traffic is set as a condition meaning a mix of manual drivers and HAVs but no other levels of automation and no complete penetration by HAD
- the HAV drives are compliant but have problems with sensor range and drive obtusely with little adjustment and compensation (late reactions and no proactive behaviour)
- the driver tends to drive less compliant but adapts and compensates better (usually early and on-time reactions and proactive acting in terms of anticipation)

9.2 Methods

The independent variable is the change in automation of different agents and functions and, thus, performance variability deviations in the system that represents the input for the model. Here, the driver and automation data is taken from Grabbe et al. (2022b), which is based on interviews, surveys, and simulator observation. This results in five different overtaking scenarios (see Figure 26). In general, each scenario represents overtaking on a rural road, including the four agents: ego vehicle (EV), lead vehicle (LV), rear vehicle (RV), and oncoming vehicle (OV). The EV is following LV and wants to overtake it, RV is following EV, and OV is driving free on the oncoming lane representing a platoon of OVs. Additionally, the scenario can be divided into five stages: follow, swerve, pass, merge, and get-in-lane. In the first scenario, only manual drivers exist. In the second and third scenarios, EV or OV are replaced by HAD, respectively. The fourth scenario combines scenarios two and three, where HAD replaces both EV and OV. In the fifth scenario, EV is removed by shared & traded control between the driver and automation following the recommendations by Grabbe et al. (2022b).

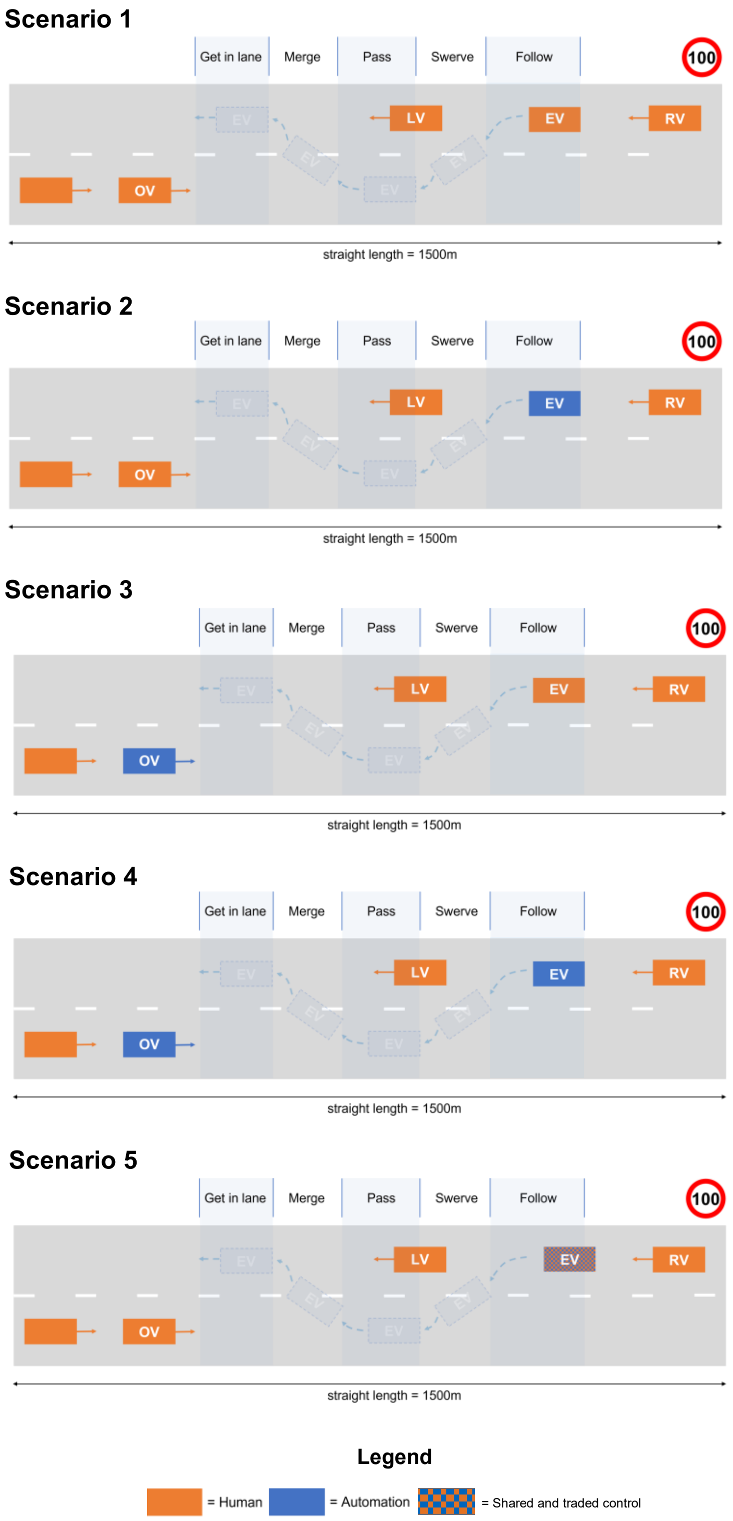


Figure 26. Overview of the differently automated overtaking scenarios.

The reasons for these five scenarios are as follows. Scenario one is the baseline, and it is interesting to compare the systemic affects and effects when changing the EV or/and OV by automation as these two agents are more critical for the successful outcome to overtake safely than LV and RV. Furthermore, it is interesting to analyse the system behaviour in the case of the shared & traded control concept, which is the system design recommendation in Grabbe et al. (2022b). However, scenario five is just a theoretical consideration as the FRAM model has to be adapted, meaning that new functions and couplings will probably arise combined with changing performance variability values.

The dependent variables depict the system/model behaviour on a global and functional level based on Grabbe et al. (2022b). On the global level, the *global system variability* (*GSV*) is used to show the accumulated variability of all functions and their interactions with the whole system for one specific condition/scenario. On the functional level, the *overall functional coupling variability* (*OFCV*) is applied to identify critical functions with high potential for functional resonance, offering functional prioritisation of their impact on the system. For example, a high value means that the function has a large systemic effect and/or is largely systemically affected, and/or a high variability accumulates in and around the function. The *OFCV* represents a complex combination of *functional variability* (*FV*) and *system resonance* (*SR*). The *FV* illustrates the variability that a function directly receives (represented by the *uplink functional coupling variability* (*ULFCV*)) and transfers (represented by the *downlink functional coupling variability* (*DLFCV*)) without considering their interaction and effect in the system sufficiently. The *SR* reflects the interaction and complexity of a function in the system, incorporating non-linearity, emergence, and dynamic of the system by weighting the system-wide impact (represented by the *weight as upstream* (*WaU*)) and affectedness (represented by the *weight as downstream* (*WaD*)) of a function.

The expected results in terms of the predictive validity of the FRAM model concerning the five scenarios are:

- differences in overall, stagewise, agentwise, and function type-wise *GSV* (different spots of high destabilisation in the system)
- differences in *OFCV*, *FV*, and *SR* for respective functions and thus different spots of high destabilisation on the functional level, whereas the differences in *SR* are less compared to *OFCV* and *FV* because the system and

the basic scenario are the same, which is why the FRAM model with its functions' and couplings' structure remains the same

- differences in patterns to strongly destabilise the system

Moreover, the aim is to identify potential differences between the scenarios based on the previously mentioned metrics to derive design recommendations in order to improve system safety.

9.3 Results

It has to be pointed out that the results are only analysed on a descriptive level, not using inferential statistics.

9.3.1 Global level

On the global level, the *GSV* is evaluated overall, stagewise, agentwise, and function type-wise. It has to be noted that the *GSV* here is a relative rather than absolute value meaning that, for example, the stagewise *GSV* shows the *GSV* in relation to the number of functions within a respective stage. We can see significant differences between the scenarios concerning the overall and the stagewise GSV (see Figure 27). Scenario five is the most stable. Instead, scenario four is the most unstable one. However, the order from stable to unstable between the five scenarios is in accordance with the expectations. In the Follow stage, scenarios two and four are significantly more unstable than the others. In the Swerve stage, scenarios three and four are significantly more unstable than the others. In the Pass and Merge stage, scenario four is significantly more unstable than the other scenarios. Scenarios one and three are significantly more unstable in the Get-in-lane stage than in the other scenarios. Furthermore, in the Swerve, Pass, and Merge stage, scenarios three and four are significantly more unstable than the other scenarios, indicating that the other agents (LV and RV), as well as OV itself, are more negatively influenced by the OV Automation and EV & OV Automation. Interestingly, the proportions between the stages are almost the same for each scenario. For example, the Pass stage shows the highest GSV, and the Get-in-lane stage shows the lowest GSV within each scenario. Thus, it should be pointed out that the Pass stage is the most critical in every scenario. Moreover, it becomes evident how the interactions and interdependencies lead to different consequences regarding the *GSV*, e.g., the automation of OV in addition to

EV's automation leads to a significant increase of *GSV* in the Pass stage. However, the single automation of either EV or OV only leads to a slight increase in *GSV* compared to scenario one. Instead, in the Follow stage, the automation of OV in addition to EV's automation leads to a significant decrease of *GSV* compared to the single automation of EV. These two examples clearly show how OV automation, in addition to EV automation, can have both a positive and negative contribution to the functional resonance in the system.

A potential for LoDA 4 can only be seen for the EV in the Merge and Get-in-Lane stages and for the EV functions related to the performance of LV, RV, and OV in the Swerve, Pass, and Merge stage. Otherwise, a shared & traded control concept for EV should be preferred. Overall, the model responses for the overall and stagewise *GSV* can be described as plausible.

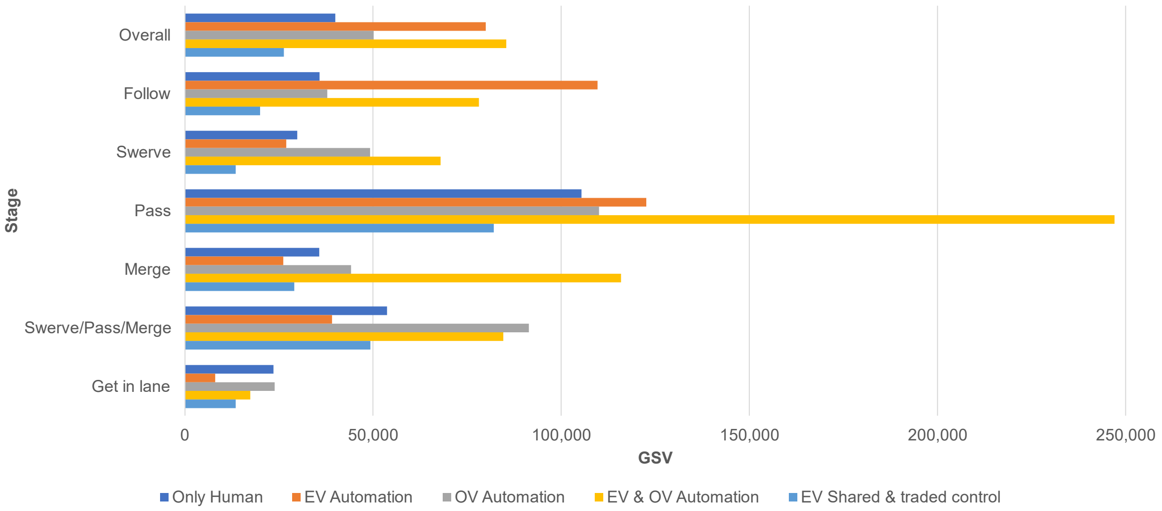


Figure 27. Comparison of the global system variability for the overall system and per stage between each scenario.

Figure 28 depicts the *GSV* per agent. For scenarios two and four, the system is highly destabilised by the EV compared to the other scenarios. Whereas, for scenarios three and four, the system is highly destabilised by the OV compared to the other scenarios. This matches the expectations. In both cases, it is noticeable that the additional automation of a second vehicle compared to a single automation leads to a slight decrease in *GSV* related to the respective agent rather than an increase. However, this positive effect is outweighed by negative impacts on the other agents. This illustrates why a systemic approach with the perspective of all involved agents is essential compared to a one-sided view for only one agent. Furthermore, the changing scenarios do not affect LV and RV stability. However, the automation of EV, EV & OV, and the shared & traded control concept for EV led to improved stability for RV. Overall, the agentwise *GSV* analysis shows plausible responses.

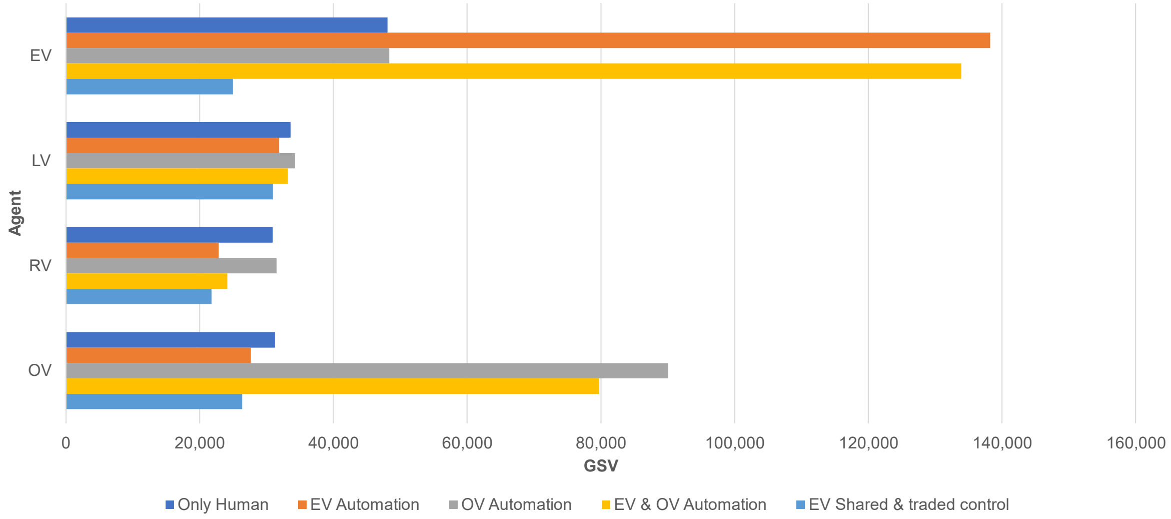


Figure 28. Comparison of the global system variability per agent between each scenario.

Significant differences between the scenarios can be found for the function type-wise GSV for cognitive and perception functions (see Figure 29). This is not surprising as automation frequently shows high performance variability outputs concerning these types of functions compared to drivers. Especially scenarios two and four have a highly destabilising character concerning perception and cognitive functions. Scenario five leads to a significantly decreased *GSV* in all four function types. The function type-wise *GSV* response behaviour comes out as plausible.

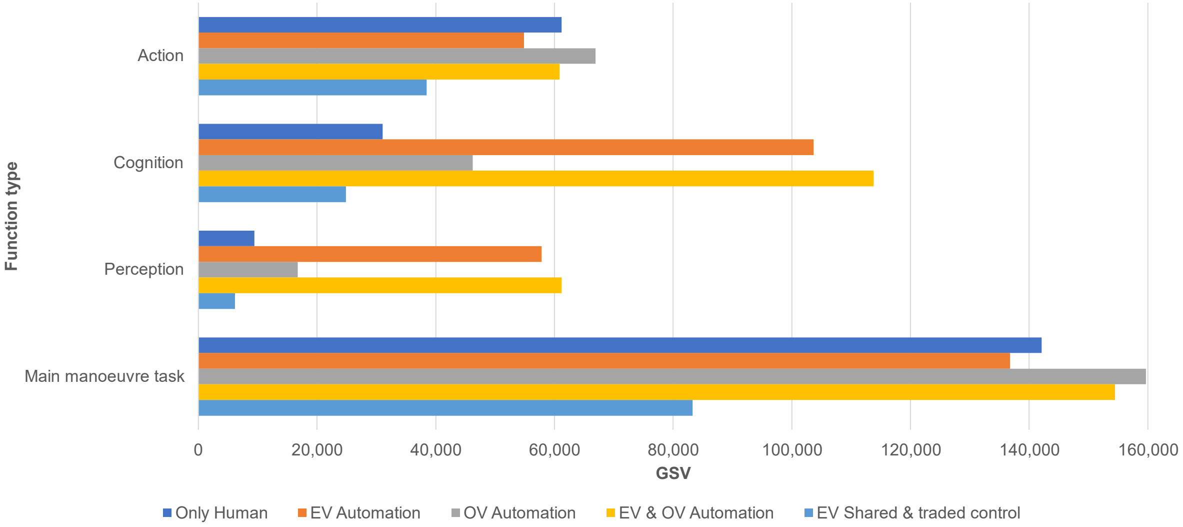


Figure 29. Comparison of the global system variability over the function types between each scenario.

9.3.2 Functional level

It is impossible to compare all 210 foreground functions in the examined FRAM model. Therefore, a selection has to be made which is based on risk or critical functions identified for each scenario. Note that this analysis defines the priority of intervention to start the investigation from the most critical functions, which is why different and further priorities and analyses are possible and even necessary. The reason is that the critical functions have the highest potential to propagate functional resonance within the system, leading to emerging events. In particular, the *OFCV* of each function was prioritised and ranked for each scenario using the scree test following the approach by Grabbe et al. (2022b). This leads to the following risk functions concerning the different agents for each scenario (see Table 2). The risk functions which represent critical paths are highlighted in green. These critical paths will be explained below in Section 9.3.3.

Table 2. Overview of risk functions per agent and scenario.

Function	Agent	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
follow LV (EV)	EV	x 🟩	x	x 🟩	x	x
maintain headway separation (EV)		x 🟩		x 🟩		
perform overtaking (EV)		x 🟩	x 🟩	x 🟩	x 🟩	x 🟩
assess opportunity to overtake safely (EV)		x	x 🟩	x	x 🟩	x
adopt overtaking position (EV)		x		x		x
keep in lane (EV)		x		x		x
assess any new info for safety of manoeuvre again (EV)		x	x	x	x	x
merge back into starting lane (EV)		x	x	x	x	x
assess situation to enter safely (EV)		x	x	x	x	x
increase speed (EV)		x		x		
assess any new info for safety of manoeuvre (EV)		x	x	x	x	x
abandon manoeuvre (EV)		x	x		x	x
glance in nearside wing-mirror (EV)		x				
observe oncoming traffic (EV)			x 🟩		x 🟩	x
check LV is not about to change speed (EV)			x		x	
assess road conditions (EV)			x		x	
assess gap ahead of LV (EV)			x		x	
recheck road ahead (EV)			x		x	
continue observing road ahead (EV)			x		x	
assess availability of safety margin in case of abort (EV)			x		x	
re-recheck road ahead (EV)			x		x	
anticipate course of LV (EV)			x		x	
assess overtaking opportunity again (EV)			x		x	
watch for hazards located at road side environment (EV)			x		x	
driving free (OV)	OV	x		x		x

keep in lane (OV)		x		x		x
respond to EV's passing problems (OV)		x		x	x	x
react to overtaking EV (OV)		x		x	x	x
recognise overtaking EV (OV)				x	x	
determine whether EV's overtaking can be safely completed (OV)				x	x	
observe for overtaking oncoming vehicles (OV)				x	x	
detect EV's swerving into oncoming lane to pass (OV)				x		
recognise EV's experiencing problems to pass (OV)				x		x
anticipate required speed adjustments (OV)				x		
follow EV (RV)	LV & RV	x	x	x	x	x
react to being passed (LV)		x	x	x	x	x
keep in lane (LV)		x	x	x	x	x
respond to EV's passing problems (LV)		x		x		x
adjust to adequate speed (LV)		x		x		x
respond to EV's passing problems (RV)		x		x		x
driving free (LV)		x		x		x
react to EV's overtaking (RV)		x		x		x
recognise EV's experiencing problems to pass (LV)		x		x		x
recognise EV's experiencing problems to pass (RV)		x		x		x

When comparing the *OFCVs* for EV’s risk functions in each scenario, different spots of high destabilisation on the functional level can be identified (see Figure 30 part A). However, if significant differences exist, two scenarios show nearly similar values in each case. Thus, it is not unique to just one scenario. There are three functions where scenarios one and three each show a significantly higher destabilising manifestation (blue-shaded areas). The function "maintain headway separation", in particular, shows huge differences. On the other hand, there are 18 functions where scenarios two and four each show a significantly higher destabilising character (orange shaded areas). Here, the functions "observe oncoming traffic" and "assess opportunity to overtake safely" stand out. When comparing the *OFCVs* for OV’s risk functions in each scenario, different spots of high destabilisation on the functional level can also be identified (see Figure 30 part B). However, large differences can only be found concerning seven functions where scenarios three and four each show a significantly higher destabilising character (grey-shaded areas). No differences can be identified concerning LV’s and RV’s risk functions. The results in terms of the *OFCV* for EV, LV, RV, and OV are plausible, especially against the background of the *GSV* results mentioned previously.

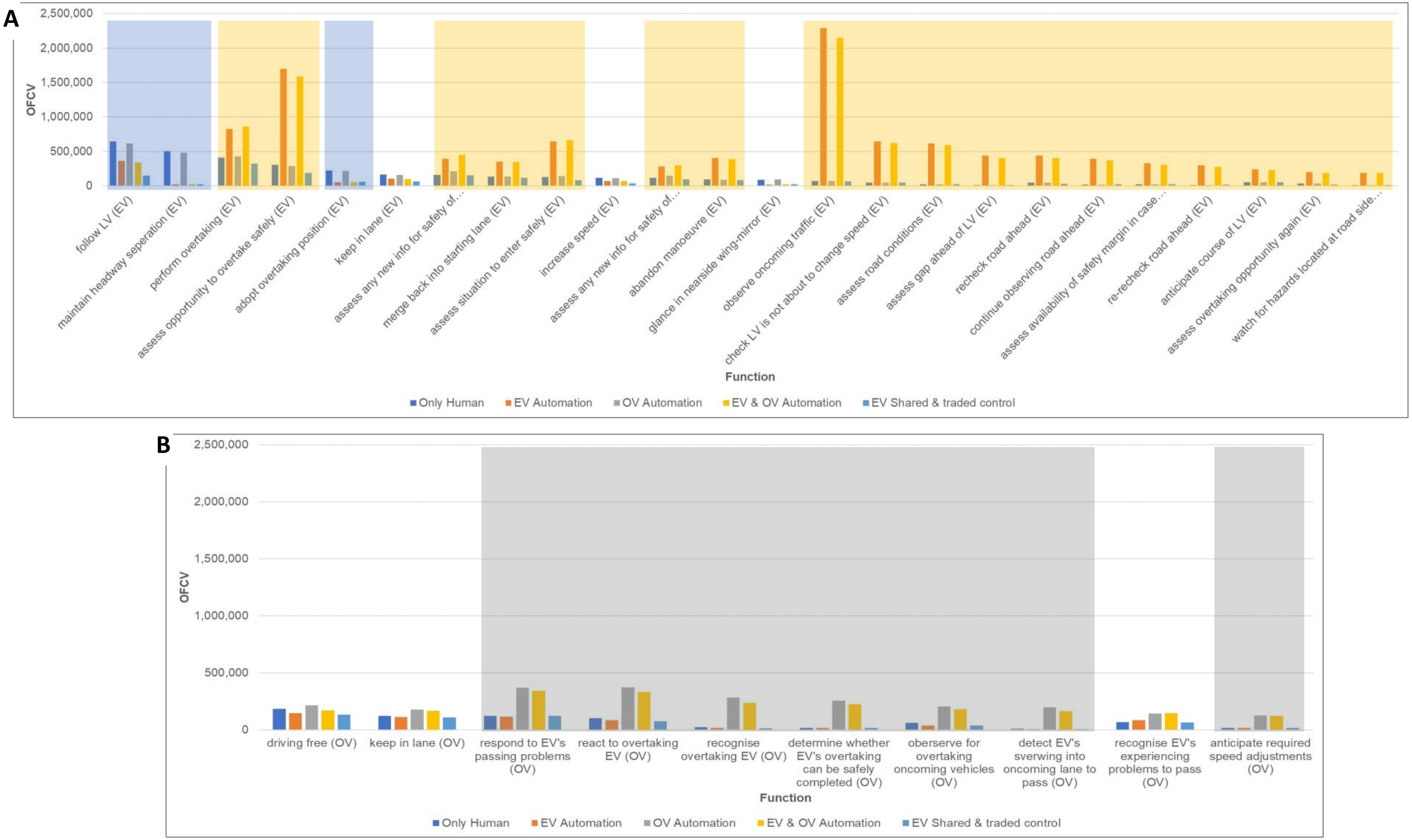


Figure 30. Risk functions for EV (a) and OV (b) based on the overall functional coupling variability between each scenario.

In the following, the *FV* and *SR* are compared in order to get a deeper understanding to distinguish the interaction and variability of risk functions as well as to consider whether these functions are rather affected than having a strong impact, vice versa, or both. Figures ***31*** and ***32*** illustrate the two dimensions *FV* (*DLFCV+ULFCV*, blue and orange stacked columns) on the left y-axis and *SR* (*WaU+WaD*, blue and orange stacked lines) on the right y-axis for the risk functions of EV and OV (x-axis), respectively, for each scenario. Furthermore, the blue, orange, and grey shaded areas (here called columns) with Roman numerals illustrate the differences between the three scenario classes of one and three, two and four, and three and four, respectively. The letters within the columns represent functions with different characters that must be distinguished.

For the risk functions of EV, it can be seen that the *SRs* are nearly the same for every scenario, as assumed. Only slight differences exist. The three highly destabilising functions (blue columns I and III) in scenarios one and three have a high *SR* and *DLFCV,* indicating their high influencing potential to propagate much variability through the system. This potential is significantly dampened in the other three scenarios. For scenarios two and four, the following applies. The functions in column II and column IV part B have a high *SR* and *ULFCV,* indicating their high potential to be affected and receive much variability. This incoming variability is significantly reduced in the other three scenarios. Instead, the functions in column IV, part A, column V, and the first function in column VI, part A have a high *SR* and *DLFCV,* indicating their high influencing potential to propagate much variability through the system. This potential is significantly dampened in the other three scenarios. Here, the function “observe oncoming traffic” also has a high ULFCV, making it highly critical. The three remaining functions in column VI, part A, have a moderate *SR* and a high *DLFCV* and *ULFCV*. Thus, they are also critical but with a reduced propagating effect through the system. This criticality is decreased in the other three scenarios. The five functions in column VI part B have a moderate *SR* and a high *ULFCV,* indicating their moderate potential to be affected and receive much variability. This incoming variability is strongly reduced in the other three scenarios. Finally, the two functions in column VI part C have a moderate to slight *SR* and a high *DLFCV*, making them variability-prone but usually remain without adverse consequences. This variability-proneness is reduced in the other three scenarios. Scenario five unifies all previously mentioned variability dampening potentials and also shows no further increased functional resonance in the three functions in the white-shaded areas compared to the other scenarios.

Regarding the risk functions of OV, it can be seen that the *SRs* are nearly the same for every scenario, as assumed. Only slight differences exist. One exception has to be made for the function "observe for overtaking oncoming vehicles (OV)" in column I, part D. Here, the SR is decreased by 40% in scenarios three and four compared to the other three scenarios. For scenarios three and four, the following applies. The functions in column I, parts A, C, and E have both a high *DLFCV* and *ULFCV*, as well as high (in case of A) to moderate *SR,* which makes the function in part A highly critical and the remaining functions in part C and E critical due to reduced system propagation. This criticality is greatly reduced in the other three scenarios. The function in column I, part D, has a high *DLFCV* and moderate *SR,* indicating its moderate influencing potential to propagate much variability through the system. This potential is significantly dampened in the other three scenarios, whereas the propagating effect is higher in the other three scenarios in the case of induced variability. The functions in column I, part B, and column II have a high *ULFCV* and moderate *SR,* indicating their medium potential to be affected and receiving much variability. This incoming variability is greatly reduced in the other three scenarios. Interestingly, a noticeable difference exists for the function "keep in lane (OV)" which was not seen in terms of the *OFCV* in Figure 30 part B. This function has a high *ULFCV* and *SR* in scenarios three and four, indicating its high potential to be affected and receive much variability. Again, this incoming variability is largely reduced in the other three scenarios. Scenarios one, two, and five unify all previously mentioned variability dampening potentials and also show no further increased functional resonance in the three functions in the white-shaded areas compared to scenarios three and four.

For the risk functions of LV and RV, it can be seen that the *SRs* are nearly the same for every scenario, as assumed. Only slight differences exist. In particular, no apparent differences can be found for the *FVs* in the five scenarios for both *DLFCV* and *ULFCV*.

Ultimately, the risk functions are considered in terms of the Functional Variability-System Resonance Matrix (FVSRM) (Grabbe et al., 2022b) to represent the criticality of functions and their potential for functional resonance (see Figure 33 and Figure 34). The following colour scheme applies: green for uncritical functions, blue for high variable functions with low system resonance, yellow for medium variable functions with medium system resonance that are between uncritical and critical functions, orange for low variable functions with high system resonance, and red for critical functions. Here, the orange and blue areas refer to functions that must be viewed cautiously due to their special features. Functions in the blue area are typically variability-prone but usually remain without adverse consequences (i.e., accidents) because they have a low systemic resonance. Functions in the orange area are functions where variability rarely occurs, but when it happens, a strong systemic effect (destabilisation of the system) and, consequently, a high probability of accidents must be expected. Moreover, it can be argued that these functions are success factors demonstrating resilience because they have little variability despite their strong affectedness and provide stability with a system-wide effect. In general, the functions in the orange area pose a greater hazard than the blue ones if performed inappropriately and are, therefore, to be assessed as more critical.

In terms of EV's risk functions and their criticality, differences between the scenarios can be seen (see Figure 33). Here, scenarios two and four show more critical functions than scenarios one and three, which are further reduced in scenario five. In addition, the criticality of several functions switches between the scenarios, for example, the function "maintain headway separation" is critical in scenarios one and three, which is reduced in the other scenarios, changing to a success factor. Nevertheless, some functions exist which are critical for all scenarios but with a different degree of functional resonance, e.g., "perform overtaking (EV)" and "merge back into starting lane (EV)". Furthermore, scenarios one, three, and five show more uncritical functions but still consist of some functions with a high *SR* despite a low *FV* posing still a risk, e.g., the function "abandon manoeuvre (EV)". Moreover, different variability-prone functions usually remaining without adverse consequences exist: "glance in nearside wing-mirror (EV)" for scenarios one and three, and "watch for hazards located at road side environment (EV)" for scenarios two and four.

Similar differences between the scenarios can be observed for the risk functions of OV and their criticality (see Figure 34). In contrast, scenarios three and four show considerably more critical functions than the others.

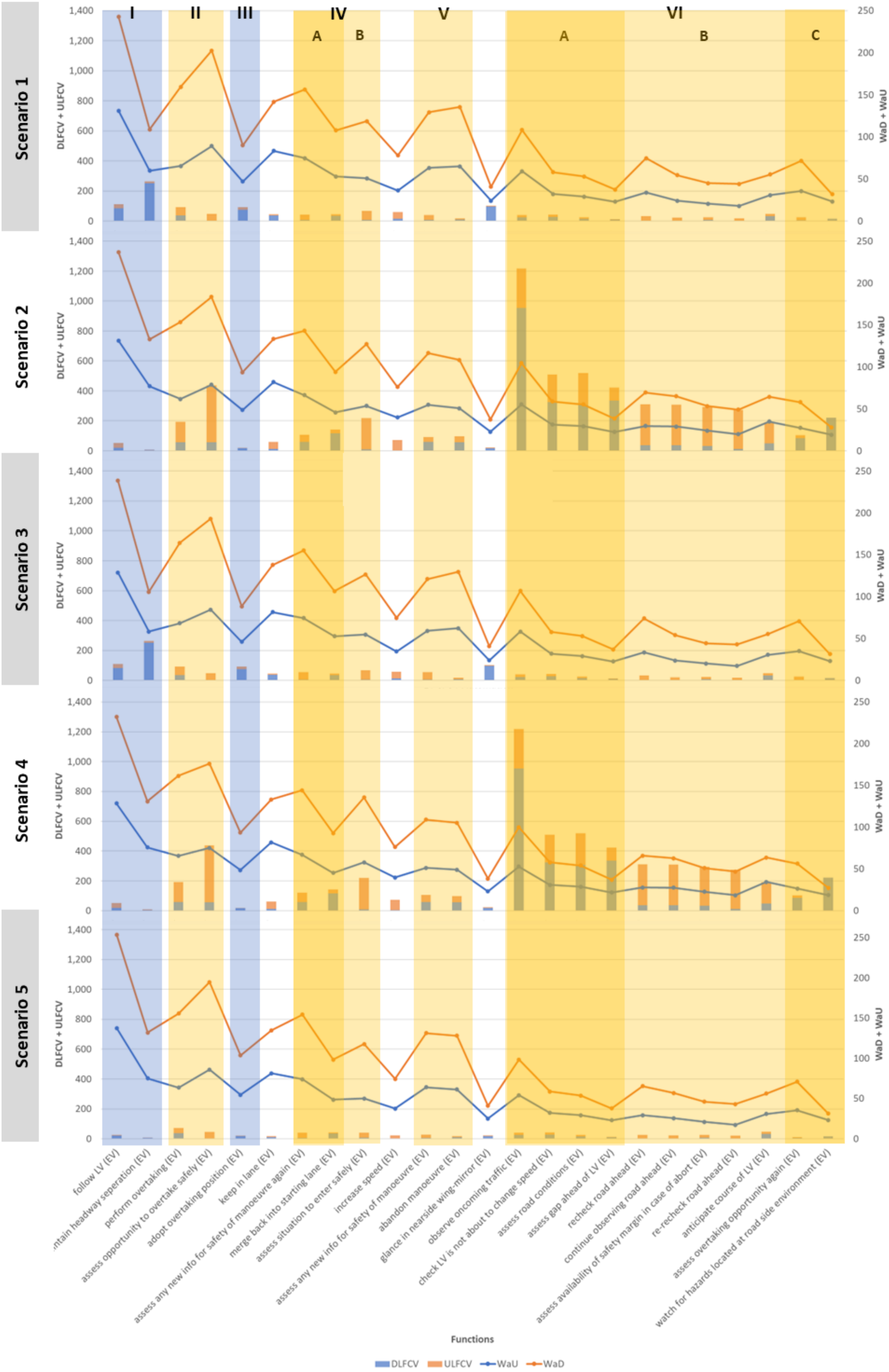


Figure 31. EV’s risk functions composed of functional variability and system resonance between each scenario.

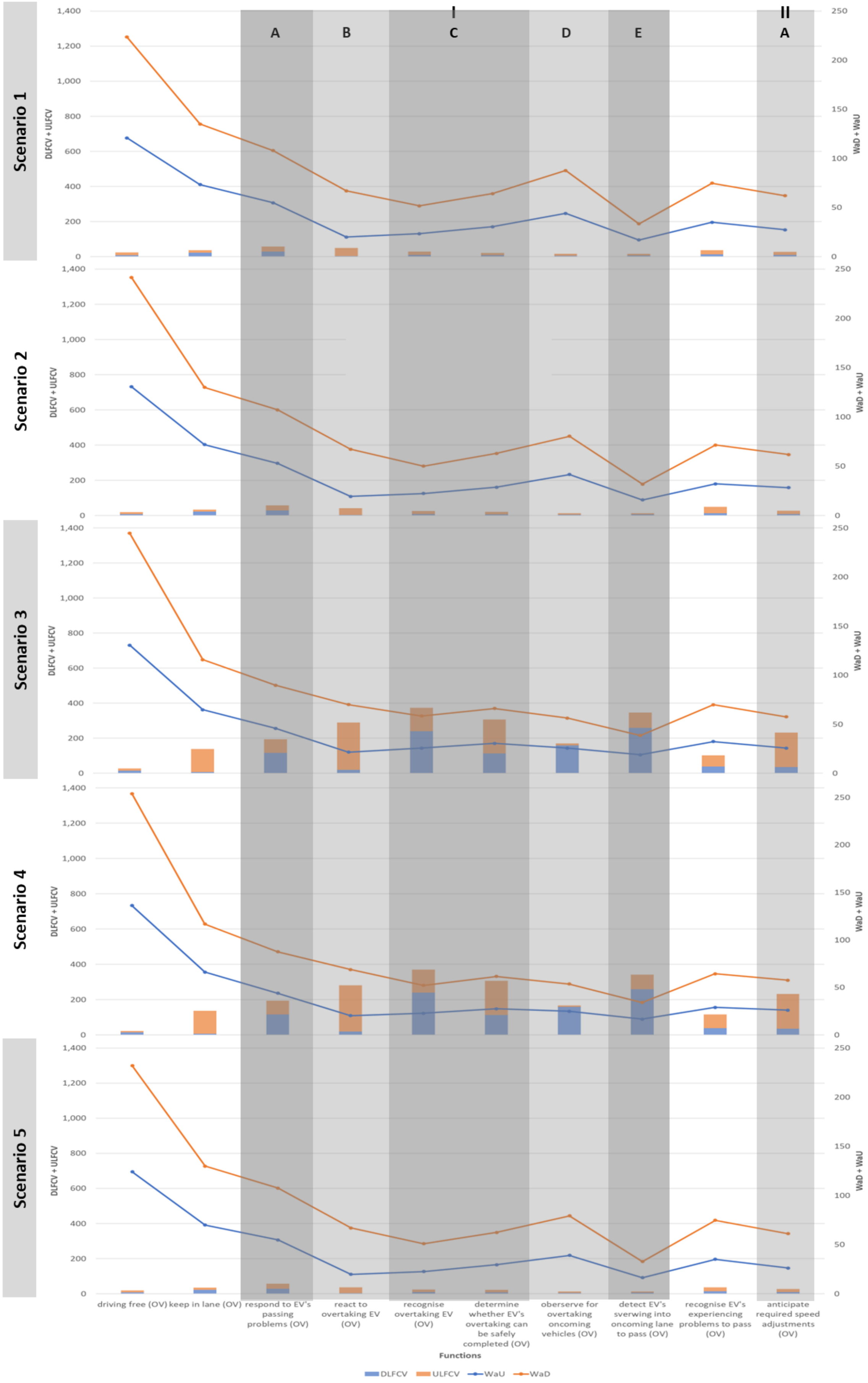


Figure 32. OV's risk functions composed of functional variability and system resonance between each scenario.

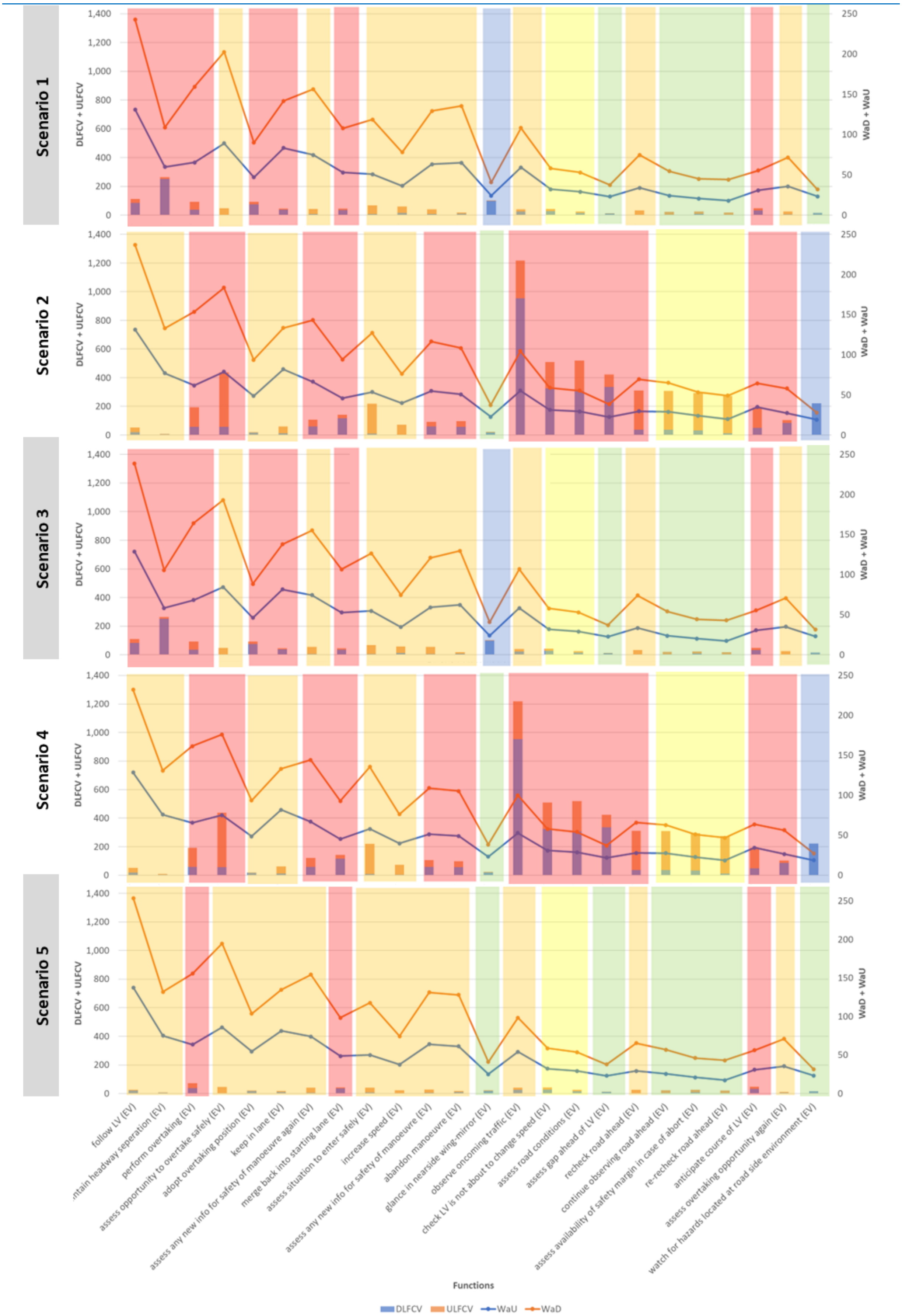


Figure 33. EV's risk functions composed of functional variability and system resonance between each scenario, representing their criticality according to the FVSRM.

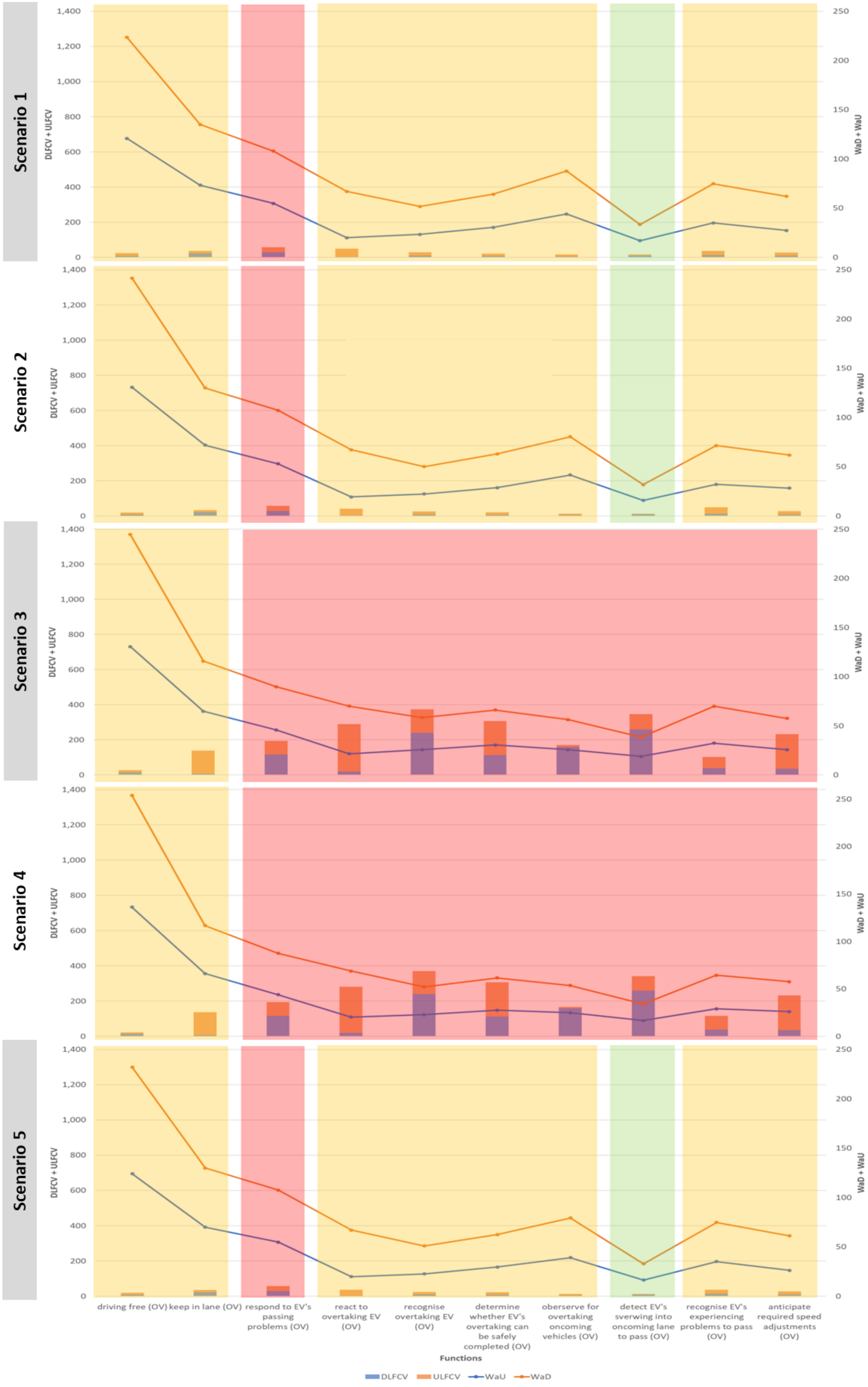


Figure 34. OV's risk functions composed of functional variability and system resonance between each scenario, representing their criticality according to the FVSRM.

9.3.3 Patterns

The quantitative evaluations on the global and functional level shown previously enabled a systematic and structured analysis even for largely complex FRAM models ensuring comprehensive results where a simple visual discussion would not be sufficient or overwhelming. In principle, the quantifications were used to obtain an overview of the influence and affectedness of system functions and their variabilities and interactions in the system in comparison between the five scenarios to identify spots with high potential of functional resonance representing leverage points for system design. Finally, this information has to be qualitatively reflected in the model to enable the patterns to be fully understood. According to Meadows (1999), leverage points are places within a complex system where small shifts or adjustments in one or a few element(s) (e.g., functions) can produce significant changes in the system and thus represent points most effectively to intervene in the system. It should be emphasised that the illustrative results do not aim to represent a complete risk or safety analysis but rather show FRAM's potential for a complex safety analysis based on selected results. In the following, this is exemplified by a maximum of three critical paths for both agents EV and OV for each scenario. Only EV and OV are considered, as there are large differences between the scenarios for these two agents. In addition, it is limited to a maximum of three critical paths each, otherwise, it goes beyond the scope. The three functions with the highest *OFCV* were selected, whereby at least a score of 250,000 had to be fulfilled. Otherwise, it does not represent a critical path with a strong destabilising system character. This results in three critical paths for scenarios one and two each, six for scenarios three and four, and one for scenario five. The risk functions representing these critical paths are highlighted in green in Table 2.

In this work, a critical path is defined as the direct couplings between a risk function and its upstream and downstream functions, so all indirect couplings are not considered to avoid confusing the analysis. Here, in the sense of the Pareto principle, the direct couplings of a risk function represent the interactions with the highest leverage to improve the resilience of the system in comparison to the indirect couplings, which also can positively contribute to improved system resilience but to a significantly lower extent such as fine-tuning. In the following, a critical path is exemplified using the risk function “follow LV (EV)” in scenario one which will be referred to in the following as function in focus (FiF) (see Figure 35). The remaining referred to in the following as function in focus (FiF) (see Figure 35). The remaining critical paths of the other risk functions can be taken from Appendix A. The FiF is highlighted in light blue (filled-in hexagon), and the upstream and downstream couplings are highlighted by orange and blue lines, respectively. Additionally, the types of functions are labelled by the respective colours: yellow for perception functions, blue for cognition functions, and green for action functions. Furthermore, the associated agents and stages are illustrated through dashed lines. Finally, every function's hexagon has coloured dots at the top left and right, and bottom left and right marking the degree of *ULFCV* and *DLFCV*, and *WaD* and *WaU*, respectively. The following colour scheme applies for the respective metric: green means lower than 5%, orange means above 5% but lower than 30%, and red means higher than 30%, according to the FVSRM.

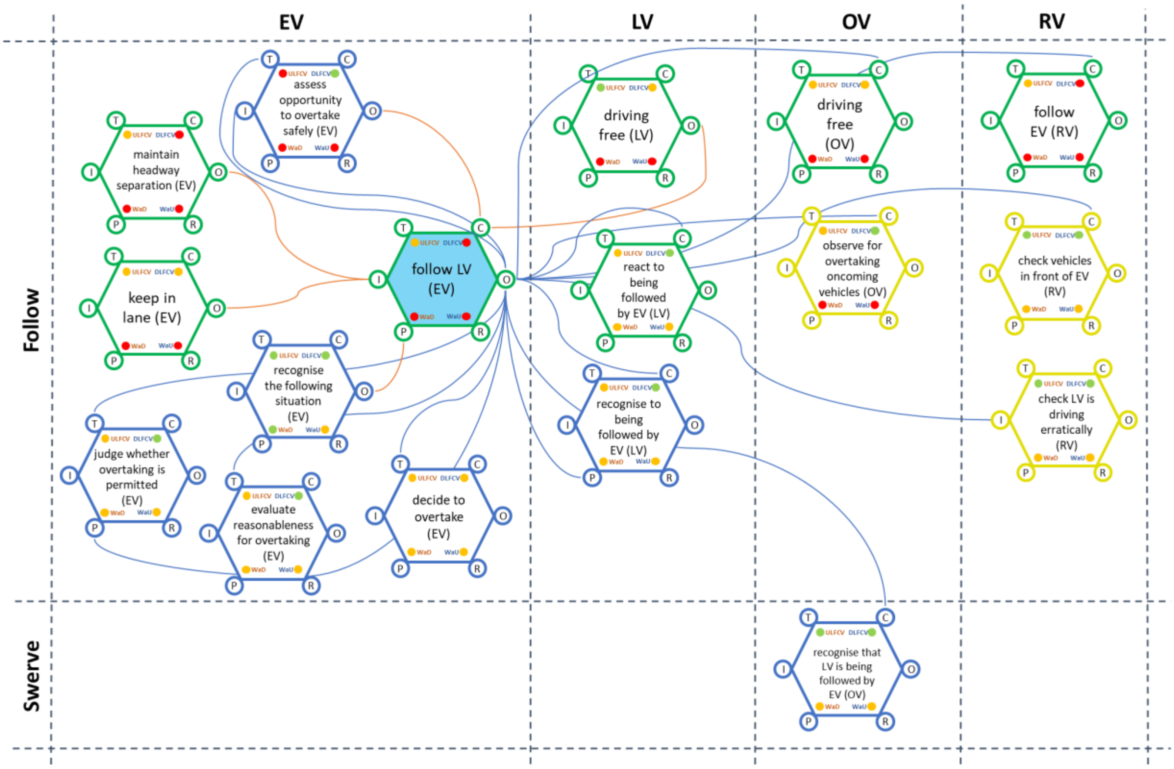

EV
LV
OV
RV
Follow
Swerve
T
C
I
O
P
R
ULFCV
DLFCV
WaD
WaU
assess opportunity to overtake safely (EV)
maintain headway separation (EV)
keep in lane (EV)
follow LV (EV)
recognise the following situation (EV)
judge whether overtaking is permitted (EV)
evaluate reasonableness for overtaking (EV)
decide to overtake (EV)
driving free (LV)
react to being followed by EV (LV)
recognise to being followed by EV (LV)
driving free (OV)
observe for overtaking oncoming vehicles (OV)
recognise that LV is being followed by EV (OV)
follow EV (RV)
check vehicles in front of EV (RV)
check LV is driving erratically (RV)

Figure 35. The critical path of the function "follow LV (EV)" in scenario one. It is the same for scenario three but with different values of the metrics *ULFCV*, *DLFCV*, *WaD,* and *WaU*.

The FiF has five uplinks with moderate incoming variability coming from four EV's functions and one LV's function solely in the Follow stage, which are three action- and two cognition functions. The incoming variability mainly comes from action functions, especially the function "maintain headway separation". Overall, the FiF is predominantly influenced by the same agent EV and slightly through the driving behaviour of LV. Moreover, the FiF has 15 downlinks transferring a high variability output, mainly in the Follow stage (14) and one function of OV in the Swerve stage, which are six cognition-, four action-, and three perception functions. It impacts its cognition functions related to the overtaking decision process forming a mutual interaction. It also influences the driving performance of the other three agents and potential anticipations by OV and RV. In particular, the downstream couplings to the functions “assess opportunity to overtake safely (EV)”, “driving free (OV)”, “observe for overtaking oncoming vehicles (OV)”, and “follow EV (RV)” are critical as the coupled downstream functions have a high *SR* indicating their high potential for functional resonance in turn. In summary, the FiF has a high potential to destabilise the system in the Follow stage.

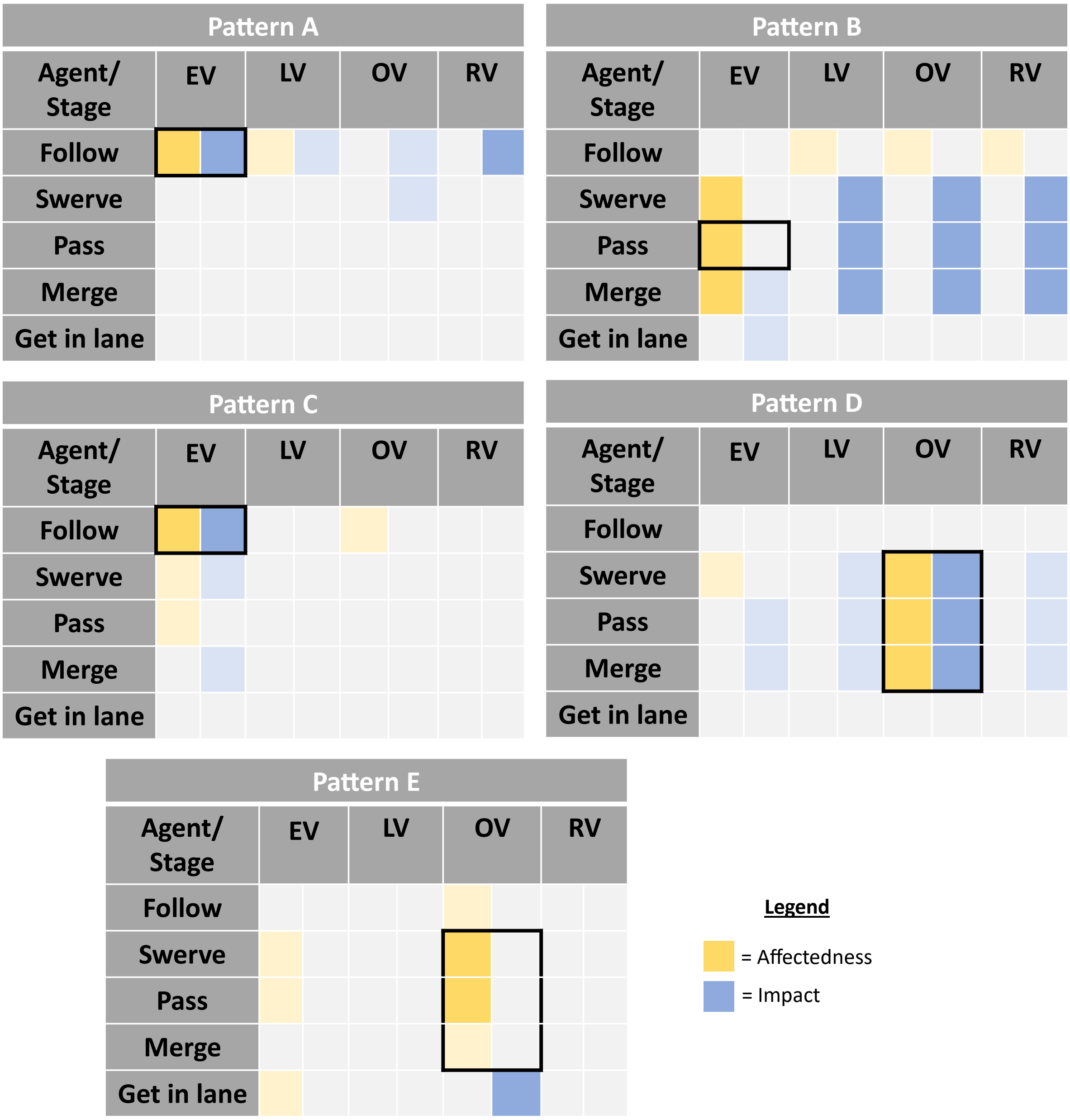


Figure 36. Overview of identified patterns between scenarios.

Based on the critical paths, five different patterns (A-E) could be identified (see Figure 36). Figure 36 shows the affectedness by upstream functions (orange cells) and the impact by downstream functions (blue cells) for the patterns on a stage and agent level. The intensity of colour represents the extent of affectedness or impact: the more intense, the higher the affect or effect. The location of the respective risk function(s) is visualised through black-framed cells. The patterns generally represent leverage points in system design to improve resilience due to their huge system affectedness and/or impact. In principle, the upstream functions can adapt to facilitate the performance of the risk function. In contrast, the downstream functions can adapt to compensate for potential bad influences of the risk function. In turn, the risk function can facilitate the performance of its downstream functions, which may provide much system stability. These patterns represent two sides of the same coin as providing a positive or negative contribution to road safety. If the adaptations are coordinated appropriately, the system works safely, and a potential accident will be avoided; if not, an accident might develop and probably occur. So, the patterns have a strong stabilising or destabilising character which depends on the fitting of performance variability. That is also the reason why an activity performed in two different contexts with the same variability can lead, in one case, to an acceptable system performance but, in the other case, to an accident. Thus, context and the interactions in the whole system matter.

Pattern A has a high interrelatedness but a low intrarelatedness. The system can be strongly destabilised in the Follow stage, based on the following behaviour of EV. It is mainly influenced by EV and slightly through LV within the Follow stage, and it impacts the overtaking decision process of EV and the driving performance of the other three agents, especially RV, within the Follow stage. It can be argued that pattern A sets the starting situation for a successful overtaking manoeuvre as it affects the information gathering concerning the overtaking decision. In particular, assessing the opportunity to overtake safely can directly affect the following process in turn, creating a tight coupling. EV itself can primarily influence the initial situation, whereas the other three agents can compensate for potential bad influences to provide stability and resilience within the Follow stage.

Pattern B has both a high interrelatedness and intrarelatedness. The system can be strongly destabilised in the Swerve, Pass, and Merge stages. This process is based on the EV's overtaking performance in passing the LV. It is mainly affected by EV in the Swerve, Pass, and Merge stages consisting of operational tasks like overtaking position preparations, swerving to the oncoming lane, passing LV or abandoning the manoeuvre, and merging into the starting lane or tactical tasks such as iterative safety checks. Moreover, it is moderately affected by the operational driving performance of the other three agents in the Follow stage. In addition, it impacts the reactive driving behaviour of the other three agents in the Swerve, Pass, and Merge stage. Thus, there is a strong interdependence between the three other agents with simultaneous affects (Follow stage) and effects (Swerve, Pass, and Merge stage), creating a high degree of “cascading process” which can be amplifying or dampening depending on the variability behaviour. Furthermore, it has a slight effect on evaluating the safety of manoeuvre in the Merge stage as well as on the overtaking completion in the Get-in-lane stage of EV. Due to the strong interdepending character of pattern B, the situation can quickly resonate during the actual overtaking process with system-wide propagations if only one agent destabilises the system due to mutual resonance between every agent. Concurrently, this pattern also has the potential of high resilience as every agent can compensate to prevent an adverse event.

Pattern C has a low interrelatedness but a moderate intrarelatedness. Here, the system can be strongly destabilised in the Follow stage, based on observing the oncoming traffic and assessing the opportunity to overtake safely by EV. It is mainly influenced by EV and slightly through OV within the Follow stage, and it only affects EV, especially information acquisition concerning the safe overtaking opportunity, and EV’s following behaviour as well as overtaking decision quality representing two spots of mutual resonance within the Follow stage. Pattern C primarily determines the overtaking decision quality, which can mainly be influenced by the EV itself and compensated only by the EV. Thus, the system resilience at this point is relatively small. Similar to pattern A, the following process can directly affect the assessment of the opportunity to overtake safely in turn, creating a tight coupling.

Pattern D has a high interrelatedness as well as intrarelatedness. The system can be strongly destabilised in the Swerve, Pass, and Merge stages, based on OV’s recognising that EV is overtaking as well as OV’s response to EV’s potential passing problems. It is predominantly affected by OV and slightly through EV within the Swerve, Pass, and Merge stages. It has an impact on EV’s iterative safety of manoeuvre checking during the overtaking process as well as LV’s, OV’s, and RV’s recognising of EV’s potential problems to pass and their reactions to it and thus critically influence the cascading process during the actual overtaking manoeuvre in pattern B. Pattern D can be seen as one counterpart of pattern B heavily influencing the overtaking performance of EV in the form of the reaction by OV.

Pattern E has moderate interrelatedness and intrarelatedness. The system can be strongly destabilised in the Get-in-lane stage, based on the reaction performance of OV to the overtaking EV. It is mainly influenced by OV within the Swerve and Pass stages and slightly through EV allocated over all stages. The only impact is on the driving-free performance of OV after the overtaking of EV is finished. Therefore, the impact is relatively small, but the affectedness is very high. A potential conflict exists between OV and LV or RV when they pass after the overtaking or for the OV alone in case of leaving the road.

Table 3. Assignment of scenarios and patterns.

Pattern	**Scenario**				
	1	**2**	**3**	**4**	**5**
A	x		x		
B	x	x	x	x	x
C		x		x	
D			x	x	
E			x	x	

Each scenario can be assigned to different patterns but also show similarities (see Table 3). It can be seen that each pattern can be assigned to two scenarios each, except pattern B which applies to every scenario. In particular, scenarios three and four show the most patterns (4), followed by scenarios one and two (2) and then scenario five (1). Furthermore, pattern A only exists for manual EV drivers, whereas pattern C-E only exists for HAVs replacing EV or OV. Scenario four is special as it exacerbates the destabilisation of the cascading process in pattern B from EV's as well as OV's perspective due to adverse resonance between OV's and EV's performance by combining pattern B and D. Instead, scenario five eases the cascading process due to decreased overall variability. Nevertheless, it still poses a risk for adverse events due to high system resonance. Also, scenarios two and three exacerbate pattern B through EV's actions and cognitive processes during overtaking and anticipations and reactions by OV, respectively. Moreover, pattern A is exacerbated in scenario three by a decreased anticipation performance of OV. As pattern B occurs in all five scenarios, the function “perform overtaking” and its up- and downstreams have the highest potential to generally dampen the functional resonance in as many scenarios as possible. However, the other patterns should also be addressed by interventions to improve the system resilience as they significantly impact system stability and resilience. Netherveless, as just a few risk functions were analysed more precisely, further important patterns still exist. However, the five presented patterns represent leverage points in the system with the highest impact on resilience.

9.4 Conclusions

In the following, the results are discussed in two terms: predictive validity of the FRAM model and design recommendations to improve system resilience and safety.

Different global and functional spots of high destabilisation in the system between the scenarios could be found. In addition, plausible differences and similarities in patterns with strong destabilising character were found. For example, differences exist in how a critical function can be strongly affected through the system and can strongly influence the system. Overall, the response mode of the model on the global and functional level and its consequent strongly destabilising patterns could not be falsified concerning plausibility. Thus, we can assume increased evidence for the value and trustworthiness of the developed FRAM model. Furthermore, the analysis has proven that it is possible to implement a function-based validation to evaluate the predictive validity of a FRAM model when using a quantification approach to falsify the model behaviour on plausibility.

From a content-wise perspective, the shared and traded control concept of EV shows the best performance in terms of the *GSV* analysis, followed by the manual driver of EV. Interestingly, the automation of OV would merely lead to a slight increase in *GSV* compared to scenarios one and five. However, the automation of EV or EV and OV would significantly destabilise the system. Therefore, automation of the entire scenario is not recommended, but individual stages may be automated in the sense of an authority transfer. A major potential for LoDA 4 can just be seen for the EV in the Merge and Get-in-Lane stages and for the functions of EV which are related to the performance of LV, RV, and OV in the Swerve, Pass, and Merge stage. Otherwise, a shared & traded control concept for EV should be preferred. Furthermore, even individual functions could be automated, e.g., the following process of EV should be supported at least by an ACC system which can also be helpful to adopt a proper overtaking position. Hence, the critical path within pattern A could be mitigated. Concurrently, the tight coupling between the following process and the cognitive evaluation process regarding a safe overtaking opportunity by EV will be decoupled, probably supporting system resilience. Concerning pattern B, the use of automation is not reasonable to support the driver of EV while overtaking except for the adoption of a proper overtaking position. However, opportunities for automation of the other agents exist either to facilitate the good overtaking performance of EV or to mitigate potential bad influences of EV's overtaking manoeuvre. For instance, LV and OV could be equipped with an L2 system ensuring a stable drive and preventing abrupt brakings or accelerations and overspeeding, which especially belongs to OV. Additionally, the RV could be equipped with an ACC system to hold a sufficient gap to the LV in case EV has to abandon the manoeuvre and merge back. Moreover, the L2 systems can ensure that the LV is maintaining or even reducing its speed while being overtaken and that the OV is braking and even evading to the road side if the EV is still on the oncoming lane within a time-critical distance. Patterns C-E do not represent critical paths for scenarios one or five. Nevertheless, these patterns should also be addressed to promote resilience if something goes wrong due to high system resonance. Regarding pattern C, automation can be used to support the driver in determining whether overtaking is reasonable and permitted, increasing the probability of a positive outcome concerning the decision of whether an opportunity to overtake is safe. In addition, the available passing time or, rather, the required space to overtake the LV independently from the OV could be highlighted using augmented reality (e.g., visualising a green carpet on the oncoming lane). However, it should be emphasised that this helps drivers to recognise where the merging process in relation to LV ends. Nevertheless, the hazard of misjudgments concerning safe completions of passing in dependency to OV still exists. Potential interventions for patterns D and E were already mentioned by pattern B.

In the future, the proposed design interventions must be analysed iteratively using the FRAM model to check potential adverse consequences due to unexpected interaction effects. In addition, more extensive what-if analyses in the form of various instantiations of the FRAM model should be executed to check how the system behaves under changing scenario conditions (e.g., weather and time pressure) or in dynamic performance changes over time (e.g., fatigue and distraction). Furthermore, as indicated in Section 9.2, the FRAM model has to be adapted for scenarios two to five as the interaction between drivers and HAVs or a driver and automation in collaboration will probably lead to new functions with new couplings and changing performance variability values. Nevertheless, the results are still usable as they show probable critical paths of functional resonance in the system when introducing automation in an overtaking scenario in mixed traffic in different constellations.

10 The way ahead - General discussion, limitations, and future work

“Every beginning has an end and every end is a new beginning.”
– Santosh Kalwar –

This chapter presents a broader analysis of the research findings, the utilisation of the FRAM method, and potential areas for future research by following the previously outlined research process. The results are integrated and discussed across three dimensions: system design and validation, method evaluation, and industrial application. The primary objective is to outline how RE, specifically FRAM, can contribute to road safety in light of the introduction of automated vehicles.

10.1 Recommendations concerning system design and validation

10.1.1 System design and validation

It has to be emphasised that the following design recommendations are based on the FRAM analysis based on performance data for driver and automation in current road traffic without already introduced HAV. Therefore, the recommendations have to be understood as potential measures that have to be iteratively tested in the future mixed traffic adapting the FRAM model.

While automating the entire overtaking scenario is not recommended, individual stages of the process may be automated through authority transfer. Specifically, in the context of the EV, drivers could handle the Follow and Pass stages, while automation could take care of the Swerve, Merge, and Get-in-lane stages. However, a more differentiated approach considering specific functions per stage is recommended for better automation design based on the compensatory design principle proposed by Fitts (1951). Therefore, a shared & traded control concept for EV, as presented by Grabbe et al. (2022b), should be preferred. It should be pointed out that the proposed concept currently does not consider any driver monitoring systems (DMS) or driver readiness, which is briefly discussed in Sections 10.1.2 and 10.1.3. Based on the proposed design concept, it is noticeable that humans mostly perform the Follow and Pass stages, while automation handles the Swerve, Merge, and Get-in-lane stages, with complete automation in the last stage. Only 12% of functions are carried out in shared control, which can take place at all three information processing or driving levels (cf. Abbink et al., 2018). It is not only essential to have a suitable human-machine interface (HMI) by displays and controls from the automation towards the human (e.g., H-Mode (Flemisch et al., 2014)) so that the driver can collaborate optimally with the automation, but also vice versa in the direction of the automation using DMS (cf. Begum, 2013) to capture the state of the driver and anticipate the driver's behaviour and actions. An application for cooperative overtaking at the tactical driving level is shown by Walch et al. (2019). Besides, drivers are responsible for most perception and cognitive functions, except for the Swerve and Get-in-lane stages. At the same time, automation carries out more action functions at the operational driving level. Two of the five main manoeuvre functions (decision to overtake and overtake manoeuvre) are primarily carried out by the driver, while the other three (following the lead vehicle, adopting overtaking position, and completing the manoeuvre) are primarily automated.

In addition, based on enhanced pattern identification in Section 9, leverage points in the system could be identified as posing the most efficient and effective way to improve the system's resilience concerning the overtaking scenario by using automation or technology support. Concerning the EV, automation should be used to support the following process, as well as the adoption of a proper overtaking position. Moreover, information regarding the reasonableness and permission for overtaking should be provided. Also, the predicted spot of merging in front of the LV independently from OV should be spatially visualised, helping drivers to prevent misjudgments concerning safe completions. It becomes evident that automation opportunities exist for the other agents to facilitate a good overtaking performance of EV or mitigate potential bad influences of EV's overtaking manoeuvre. In the case of the former intention, a stable drive of LV and OV must be ensured to prevent abrupt brakings or accelerations and overspeeding. In the sense of the latter purpose, automation should be used to ensure that the LV is maintaining or even reducing its speed while being overtaken, the OV is braking and even evading to the road side if the EV is still on the oncoming lane within a time critical distance, and the RV is holding a sufficient gap to LV for the case if EV has to abandon the manoeuvre and merging back.

Grabbe et al. (2022b) compared the FRAM analysis results with a focus on the EV, represented as positive and negative contributions of the driver and automation to system safety, with state-of-the-art knowledge regarding this issue. The comparison predominantly focused on the negative contributions of the driver, as past data analyses have focused mainly on the negative contributions of drivers, such as rare or critical accidents (Bengler et al., 2017). That is also why little or no knowledge about the positive contributions of drivers exists. Furthermore, no comparison can be made for automation because LoDA 4 vehicles have not been approved yet, and data collected during test drives in California is analysed on an abstract level without a specific task-level analysis that would be required.

Grabbe et al. (2022b) concluded that the FRAM model reflects the most common causes and contributing factors of overtaking accidents by drivers. However, some of these known accident black spots cannot currently be improved by the automation of EV. However, as mentioned above, the automation of specific functions of other agents can enhance the system's overall resilience during overtaking scenarios, potentially mitigating or preventing the potential functional resonance caused by EV’s impact.

Overall, the results and system design recommendations in this research provide the following new insights concerning overtaking safety in road traffic compared to previous and less system-oriented approaches in the past (e.g., Hegemann et al., 2005; Näätänen & Summala, 1976; Reichart, 2000; Summala & Näätänen, 1988):

- Positive contributions of drivers and automation indicating spots of system resilience instead of the sole focus on negative contributions of drivers
- Additional automation risks by its negative contribution rather than mere automation benefits
- Leverage points in overtaking most effective to intervene in the system by automation interventions
- Systemic perspective including all agents and their interactions rather than just the EV in isolation
- Functional perspective rather than a structural component one for both to support decision-makers through a systemic function allocation between drivers and automation and to not predetermine any design or structure of a physical system

In terms of validation, the following applies. With the assumption of automating the whole scenario and its associated functions, particular attention should be paid to the risk functions for automation (see Table 2). This especially applies to functions of EV in the Follow and Pass stages, as well as those declared perceptual and cognitive tasks. In addition, the validation focus can be expanded to include the critical functions in the red and orange areas of the FVSRM in scenarios including automation. The validation process can likely be reduced to testing these functions, such as criteria for exclusion, to reduce the test effort. This has to be fulfilled by HAVs. Otherwise, we do not even need to carry out further tests. Further, the function allocation concept by Grabbe et al. (2022b) could be used to validate merely the functions in which automation is responsible alone or together with humans, and thus, in turn, reduce the validation effort to a level similar to the current ADASs of LoDA 2 vehicles, where humans are entirely responsible for the safety of the driving task. The only difference is that humans are not responsible for all functions, but only those allocated to them, and thus, automation takes responsibility for several other functions. The results for the validation process of HAD reveal insights for the potential reduction of test effort in two ways: First, by assuming full automation, the identified risk functions for automation can be used as criteria for exclusion. Secondly, allocating functions between humans and automation can limit the validation process to the functions allocated to automation alone or combined with the driver. Based on safety-II and RE analysis, this perspective shift offers new opportunities for resolving the approval trap.

10.1.2 Criticism of level of driving automation

The presented design recommendations for function allocation between driver and automation should be seen as a JCS (i.e., the driver-automation-vehicle system) that regards human and machine as equal partners collaborating in the sense of a human-machine coagency "by shifting the focus from human and machine as two separate units to the JCS as a single unit" (Hollnagel & Woods, 2005 p. 67). This coagency is viewed in terms of function-centeredness (Hollnagel, 2006b), where system functions of the EV needed to accomplish the overtaking manoeuvre are distributed between the driver and/or the automation, taking into account the interactions and dynamics in the system (reflected by system resonance) and the functional variabilities. In terms of SAE J3016, the resulting concept may even be realised as a highly assisted driving system instead of automated driving.

As mentioned earlier, full automation of overtaking scenarios is currently unrealistic and inadvisable as a general concept. Instead, humans must be actively involved in the driving task to a certain degree, especially perception and cognitive functions, until reliable full automation is achieved. This was also reported by Zhang et al. (2021), suggesting exploring other opportunities and roles for drivers, such as a "commander role" at strategic and tactical levels (e.g., Franz et al., 2012; Kauer et al., 2010; Walch et al. 2016, 2019), rather than limiting their role to that of a passenger or fallback operator. This approach is consistent with the design and effect space of shared control and human-machine cooperation conceptualised by Flemisch et al. (2019) or the multi-level cooperation proposed by Pacaux-Lemoine & Flemisch (2019). However, teleoperated driving or remote assistance/driving (SAEJ3016:2021) is not helpful concerning this issue since a remote operator would only intervene if the vehicle had already transferred itself to a safe state (minimal risk condition) or issued a takeover request with sufficient lead time. In fact, the active and continuous integration of a remote operator into the driving task in collaboration with the automation to supervise and perform the dynamic driving task would be necessary, which is, however, currently not planned and does not appear to be realistic or advisable because of the challenges of teleoperation (e.g., telepresence, latency, situation awareness (Tener & Lanir, 2022).

Therefore, the short- and mid-term strategy for automation to improve traffic safety in overtaking scenarios on rural roads should be to adopt the JCS approach for the traffic system (Inagaki, 2010) to realise driver-automation collaboration and coagency throughout the driving scenario to achieve their common goal of safe overtaking. Thus, a differentiated, function-centric approach must be taken, where the functions of the JCS are divided according to different function types (Parasuraman et al., 2000). Then functions are assigned to the agents, based on an FRAM analysis, in the sense of “who does what and how”. This contrasts the six rigid LoDA of SAE and its “all or nothing” approach, and as a design decision for automation, prefers considering the ten LoA according to Sheridan (1992) combined with the four function types of Parasuraman et al. (2000). This is also in line with Inagaki and Sheridan's (2019) criticism of SAE's LoDA definition, especially conditional driving automation. A further extension of the LoDA of SAE is suggested by Steckhan et al. (2022) to include optional driver interventions in the form of decisions on movement and dynamics in terms of driving parameters and driving maneuvers to fulfill the currently underestimated functional purpose of driver satisfaction besides safety, time efficiency, and ecology.

Nonetheless, at the current stage, the design recommendations in the form of a function allocation are rather static, and the agents’ roles do not change from one occasion to another or in different scenario conditions. This does not fit the actual system behaviour perfectly since technological advancements cause performance changes in the functions performed by drivers in cooperation with automation, leading to adaptations and dynamics. These adaptations sometimes result in adverse effects, such as the out-of-the-loop performance problem, loss of situational awareness, complacency or overtrust, or automation surprises (e.g., Endsley & Kiris, 1995; Inagaki & Stahre, 2004; Parasuraman & Riley, 1997; Sarter et al., 1997; Wickens, 1995). An example of this phenomenon can be seen when better brakes are introduced in vehicles to enhance road safety. Falsely assuming that the driver will continue driving in the same way, actually led to a change in driver behavior, where the driver tends to drive faster since he or she can now brake harder (Hollnagel & Woods, 2005). This effect can be explained by the concept of risk homeostasis proposed by Wilde (1982).

A more suitable approach for the future would be to implement an adaptive automation system (Inagaki , 2003) or function-congruence (Hollnagel, 2018a), which determines "who does what, how and when," and allows functions to be shared or traded between humans and automation in response to changes in situations or human performance (Inagaki, 2003) also considering the instantaneous driver status or availability. This is even more important as safety is not a permanent system condition but rather what a system continuously does. In fact, it is a dynamic global state that emerged by locally interacting and adapting elements. Each agent requires flexibility or adaptation to fit and optimise the global system dynamics. However, it must be taken into account that drivers are often not well trained, and such a complex function allocation could cause confusion despite its benefits. Furthermore, too much fragmentation of functions may not be sensible since individual functions must be carried out as a whole, sometimes by a single agent who has been well trained, as too little information or ineffective and inefficient transfer of information can happen at the interface between drivers and automation.

10.1.3 Research outlook

In future research, following an adaptive automation approach, the FRAM model for the overtaking scenario and the current design recommendations should be checked by extended “what-if analyses” (MacKinnon et al., 2021; Hill et al., 2020) in the form of various instantiations of the FRAM model under changing scenario conditions (e.g., weather and time pressure) on the one hand, and on the other hand for dynamic performance changes over time (e.g., fatigue and distraction), such as by

Hirose et al. (2021). Furthermore, not only the performance variability that can change but also new functions will emerge through the collaboration between humans and automation, which is why adapting the FRAM model in relation to the context conditions is necessary. For this purpose, the performance indicators per function must also be recalculated for the system with the new allocation of functions and iteratively adjusted because of the effect of contextual factors. In particular, the FRAM model has to be explicitly extended by the effect of performance-shaping factors due to the scenario conditions. In particular, behavioural adaptations of the driver in response to automation (e.g., Feldhütter, 2021; Ma & Zhang, 2022; OECD, 1990; Preuk et al., 2016) have to be analysed more systemically by the FRAM model. Overall, the current design concept fits the basic scenario analysed well and is a good starting point. However, it is not generally applicable and must be adapted in further iterative analyses, both in theory and practice.

10.2 Method evaluation – Benefits and limitations

10.2.1 Benefits

FRAM is a generic, agnostic method-sine-model that allows for straightforward augmentation and changes in granularity without limitations to apply and use the method of how to model a system, making it easy for users to modify without starting from scratch. This implies that the method does not propose any specific assumptions about the design or structure of the system being studied and the potential causes and relationships between causes and consequences (Hollnagel, 2012a). The "openness" and flexibility also provide opportunities for combining FRAM analysis with other tools and approaches (e.g., agent-based modelling, Petri Nets, system dynamics) for expansion and differentiation (Patriarca et al., 2018; Patriarca et al. 2017a,b; Tian et al. 2016, among others), enabling analysis of specific problems while maintaining an overall socio-technical system perspective (Ferreira & Canas, 2019). Therefore, FRAM can be used for different purposes and from different perspectives, which results in asking different questions to gain additional insights into how a complex system works depending on the aim of analysis and the epistemological positions modellers take (Sujan et al., 2023). However, it has to be assured that such combined approaches do not fall back into the traps of reductionist mathematical assumptions (Patriarca et al., 2020). Overall, a FRAM model is intended as a "toy-model" facilitating to model a system as WAD instead of WAI, which results from complex adaptation and interaction processes in the system caused by pursuing success in a knowledge and resource-constrained, goal-conflicted world (Vaughan 1996; Woods et al. 2010).

In addition, the model produced can be visualised graphically, providing a "map" of the relationships between input and output and how inputs are transformed into specific outputs, rather than just presenting the inputs and outputs themselves. A FRAM model can be seen as a white-box model to understand the inner workings of a system rather than a black-box model to focus on the outcome of the input-output relation. Therefore, qualitative and quantitative data can be integrated into one model. A benefit of using a graphical representation is that the human brain can quickly comprehend and identify patterns. Moreover, a FRAM model captures multiple n:1 rather than 1:1 couplings, also discerning their quality by using different aspects.

Furthermore, practitioners have access to various guidance materials, including Hollnagel's (2012a) book on the fundamental theory of FRAM, a concise guide on how to use FRAM (Hollnagel, 2018b), and a practical handbook (Hollnagel et al., 2014). Additionally, the use of FRAM is facilitated by software such as basic tools like FMV (Hill & Hollnagel, 2016) and Functional Model Interpreter (FMI) (Hollnagel, 2020a) or advanced pre- and post-processing tools like myFRAM (Patriarca et al., 2017c) and DynaFRAM (Salehi et al., 2021a) which promotes standardisation, systematic implementation, and structured analysis.

In addition to the original qualitative approach of FRAM, a quantitative risk or safety assessment is possible (e.g., Falegnami et al., 2020; Grabbe et al., 2022b; Hirose & Sawaragi, 2020; Patriarca et al., 2017b) integrating qualitative and quantitative data, allowing for results to be presented more easily to a wider audience. While quantification is not mandatory or recommended (Hollnagel, 2012a, p. 94), it can aid with specific issues such as promoting more reliable and valid “answers” derived by the analysis once a FRAM model was built, particularly in interpreting FRAM models related to large-scale complex systems such as the road system by preventing overwhelming qualitative representations, as noted by Ferreira & Canas (2019). Nevertheless, the quantitative results are relative rather than absolute metrics and thus represent indicators of where to look and must be reviewed qualitatively by the model, requiring careful interrogation to comprehend and anticipate potential useful interventions.

In addition to the previous benefits, FRAM and its approach in this work offer new features and unique benefits compared to other methods and models used in safety research and practice concerning the road system. In an abstract and general sense, these are the following:

- First, creating a semantically rich network or multi-digraph of functions, considering all agents and their interactions required to achieve a system goal exemplified by the overtaking scenario. This functional approach explicitly represents how the system performance, e.g., safety, emerges due to complex interactions of variabilities, which shows what a system does rather than what it is, as illustrated by structural component representations.
- Second, focussing the process and its complexity, including non-linear, dynamic interactions resulting in identifying emergentisms and paths of functional resonance on understanding how systems can be quickly stabilised or destabilised over time and space due to potentially cascading variabilities transitioning the global system state. This phenomenon is not measurable per se and is often assessed as a surprise in case of unwanted outcomes such as accidents in particular. The FRAM model visualises where and how multiple, mixed weak signals in a system can evolve into a strong signal meaning that local optimisations (locally acceptable performance) do not lead to global optimisation (globally unacceptable performance) due to adverse interactions. Therefore, it becomes possible to intervene at the right spots in the system, i.e., functions and especially their couplings, to improve overall system performance. However, usually, the focus is solely on the strong signals or outcomes themselves, in particular unwanted ones, rather than the process behind it (e.g., FTA, Bayesian networks, or the traffic simulation software PTV “Verkehr In Städten – SimulationsModell” (VISSIM)) which can predict outcomes or effects comparing different scenarios or system designs but unable to comprehend why such outcomes occur which does not help to understand and improve overall system performance especially if these predicted outcomes are negative.
- Third, indicating spots of resilience or vulnerability represented as leverage points or tipping points, which can be stabilising or destabilising depending on the performance variabilities. These spots can be considered for

interventions to better cope with unforeseen events or system behaviour, which improve the systems' adaptive capacity.

- Fourth, considering success as well as failure with a particular emphasis on success as generic insights significantly improves the learning rate because the rate at which successes occur is much higher than the rate at which failures or accidents occur.
- Fifth, providing a design space or function allocation acknowledging complexity, which anticipates the impact of new technology involving human factors at the beginning rather than the tail of the research and development (R&D) process, which often is the usual approach. Thereby, this thesis shows how FRAM can be used for a systemic function allocation between humans and automation, considering the interactions and complex dynamics of functional variabilities in a space-time continuum within and between agents in the system based on an enhancement of quantitative outputs.

This directly leads to the features related to improvements in the FRAM methodology itself:

- First, the introduced Space-Time/Agency framework combined with different function types according to the levels of information processing makes it possible to structure the system following different dimensions of analysis with different resolutions and perspectives which, according to Rasmussen & Lind (1981), makes it easier to analyse the inherent complexity in STSs effectively. Similar approaches are shown by the Abstraction/Agency framework (Patriarca et al., 2017a) or the JCS framework (Adriaensen et al., 2022).
- Second, new metrics concerning complexity and interaction were derived, making the notions of couplings and complexity to characterise STSs by Perrow (1984) more explicit or quantifiable. In addition, these metrics were intertwined with metrics for variability rather than set in isolation, resulting in *OFCV* and *GSV* comprehending complex and emergent behaviour representing safety as a system property that emerges from how elements interact and fit together (cf. Ackoff, 1971).
- Third, these new metrics allow us to calculate a global system variability to compare the stability or resilience and adaptive capacity of different scenarios or system designs, which provides quantitative indicators for safety as a positive

meaning in terms of safety-II and RE. Furthermore, the separation of *FV* and *SR* enables a weighting factor of a function's variability, making it possible to define tipping or leverage points in the system, as mentioned previously.

Overall, the approach of FRAM in this work makes a closer step to explicitly show the critical paths in a system that potentially lead to functional resonance, i.e., the adverse combination of the “regular” performance variability of multiple functions over time and over space which is the primary goal of FRAM (Hollnagel, 2012, p.8).

10.2.2 Limitations and required improvements

FRAM is an elaborate method requiring extensive knowledge of the domain and human factors (Hollnagel & Speziali, 2008) for modeling and analysis. Even with a simple model, a significant amount of time is required. More complex models and empirical data collection will require even more time and resources. Adriaensen et al. (2019) suggest limiting the model's scope to essential investigation questions to make it manageable.

Furthermore, when using FRAM to analyze highly complex systems, the graphical representation can become overwhelming and difficult to interpret due to its messy appearance, like “spaghetti models”. However, it can still provide insight into potential system dynamics. As mentioned, quantitative approaches can overcome this limitation to ensure a systematic and structured analysis.

One of the more critical limitations of FRAM is identifying system functions and their interdependencies, as well as their variabilities. Currently, these functions and variabilities are identified by studying reports, procedures, design specifications, storytelling, and conducting field observations or interviews. An experienced team of experts is required to analyse and model the system (Accou and Reniers 2019; Jensen & Aven 2018; Pereira 2013), where the quality of the output in FRAM directly depends on the team of experts and the information they provide as input for the functions and their variability (Salehi et al. 2021b). Some practical guidance material exists in Hollnagel et al. (2014), but no explicit standard for determining how much information should be included in the analytical process to define the objective, scope, and granularity of the model, as highlighted by Anvarifar et al. (2017), Grabbe et al. (2020b), Li et al. (2019), and Patriarca et al. (2017a). It is obvious that a FRAM model cannot be declared as right or wrong per se due to multiple purposes and dimensions.

However, it would be necessary to demonstrate formally how different strategies can be used for various purposes and perspectives to ensure a more guided and structured modelling process. It is apparent that the flexibility of FRAM is a blessing, as mentioned above, and a curse at the same time. Due to the low limitations or regulations regarding modelling and the strong dependency between model outputs and the competence of the modeller team, a subjective component ultimately plays a significant role in a FRAM model. Modelers have to “work” with the method and its fundamental principles, making some reasonable adaptations depending on the objective and context of application – a procedure according to a pattern or taxonomy does not exist. Using mixed methods and multiple data sources can help to ease this issue by integrating multiple limited perspectives and dimensions, adhering to verification strategies such as those proposed by Creswell & Miller (2000), and complying with the four qualitative terms of credibility, transferability, dependability, and confirmability, as suggested by Anfara et al. (2002). This can ultimately improve the quality of the FRAM model.

Nevertheless, identifying functions and their variability, particularly perceptual and cognitive processes, must be improved in further research, which must include more objective and empirical measures. In terms of road safety, researchers may need to use specific interview techniques (such as card-sorting or cognitive walkthroughs), eye-tracking methods (cf. Arenius, 2017), or a neuro-ergonomics approach (Parasuraman, 2011). Especially, data from sensor technologies can support the traditional qualitative inputs concerning the following aspects: temporal resolution, gradual differences, time-stamped data, and continuous recording, coverage, and calibration (Arenius, 2017). In addition, existing literature, such as the Driver Performance Data Book (Henderson et al., 1987), can be referred to identify sources of variability. In contrast, these approaches cannot be used for automation, so experts' assessments are currently the only option. This is due to the lack of publicly available data, although it is generally easier to identify cognitive functions in automation due to the physical architecture of software and hardware. However, using deep learning and its self-learning algorithms would hinder understanding. As a result, there is a discrepancy between WAD and WAI, as generating a model for WAD is almost feasible for drivers but remains challenging for automation. Ultimately, in this research work, a combination of literature, driving simulator studies, and interviews, each addressing different dimensions, has proven successful in integrating WAI and WAD. It should be mentioned that while driving simulators are useful for assessing operational action functions like lane-keeping or maintaining safe distances, it is challenging to assess perception and cognitive functions even with the help of eye-tracking. Structured interviews are more suitable for this, but humans' limited self-awareness and potential biases about their performance can limit the effectiveness of this approach. Overall, using multiple data to integrate these multiple limited perspectives as a mixture of qualitative and quantitative inputs is recommended. In the future, it could also be interesting to use cross-linked driving simulator studies to explicitly observe the interactions between multiple drivers, automation, and/or joint driver-automation and their resulting variabilities and adaptations within one simulation.

Lastly, validating a FRAM model is a challenging issue and ongoing concern. In line with that, Grabbe et al. (2022a) developed a framework that provides a good foundation to evaluate and increase the reliability and validity of an FRAM model, especially helping analysts to assess the cost-effectiveness of FRAM. The authors emphasised that validation in terms of FRAM is always model-individual, gradual, the result of a negotiation process, and continuous and iterative, meaning that the validity of a FRAM model and its analysis is relative rather than absolute. In particular, they distinguish between two purposes of the FRAM method, an analytic and an evaluative one which are addressed by different types of validity. Further, they stated that predictive validity is the highest maxim of validity concerning FRAM. The conclusion was that the validity and usefulness of the FRAM model by Grabbe et al. (2022b) are limited due to low specificity despite high sensitivity. However, it can be argued that sensitivity is more important than specificity in terms of FRAM because the main intention of the evaluative part of FRAM is to predict performance variability and its potential resonance. Thus, the consequence of missing a performance variability effect is significant as this could be an overlooked, crucial success or risk factor compared to the minor consequences of having false positive predictions. Nevertheless, a FRAM model should strive for an appropriate balance between both.

Besides, methodological issues exist to prove the predictive validity of a FRAM model. The main reason is that FRAM has an inherent, at least partly tautological character meaning that model results are only partly falsifiable for two reasons: interacting variables (i.e., functions) difficult to prove empirically, and no measurability of single absolute final outputs but multiple relative outputs. Thus, Grabbe et al. (2022a) concluded that an FRAM model could rather be calibrated than validated, meaning that a few interesting functions (e.g., the critical path of functions representing leverage points for system design) in the model are selected to refine their modelling for a better understanding of their potential affectedness and effects in the system with regard to specific system conditions. However, the results of the pure function-based validation approach opposed to a mere outcome-based validation approach, in Section 9, show that increased evidence for the value and trustworthiness of the developed FRAM model can be assumed. As a result, while the approach is useful for improving the fundamental understanding of system patterns as elucidated by the FRAM model, it is unsuitable for making conclusive judgments regarding the safety certification of designs in critical systems.

10.2.3 Research-practice gap

The benefits and limitations of FRAM and its application and advancement in this thesis must be reflected according to the research-practice gap (RPG).

There is a disparity between research and practice in applying systemic models and methods, particularly FRAM, as Underwood & Waterson (2012) noted. While researchers are utilising systemic methods per the latest advancements, practitioners tend to favor more traditional, linear methods that are easier to use or more popular, despite their recognised limitations, as Grabbe et al. (2022b) pointed out. Also, in everyday practice, the efficiency of a method often outweighs the drawback of reduced thoroughness (Hollnagel, 2009a p.132), which probably results in “probative blindness” (Rae & Alexander, 2017), i.e., a safety activity is believed as effective providing stakeholders with subjective confidence in safety while it does not provide the actual knowledge about real problems. This can lead to false assurance about the result of a safety analysis which may further lead to erroneous decisions. Frequently mentioned reasons for the RPG are a difficult, resource and time-consuming application (Salmon et al., 2022), reduced model validation and usability, and a potential analyst bias (Underwood & Waterson, 2012). Given these circumstances, it is crucial to establish a correlation between validation outcomes and usability, weighing FRAM's cost-effectiveness to assess its overall usefulness (cf. Stanton & Young, 2003).

The effectiveness hereby represents a trade-off consisting of the validity of the FRAM model to explain performance variability in a system and the output value of such a model represented as the potential of insights and knowledge gain. The costs are related to the resources and time used by the method. FRAM has high costs since the model development by function identification and variability data collection is time-

and resource-consuming (see Section 10.2.2). However, these can be outweighed by the high potential for new and unique insights (see Section 10.2.1) which may save considerable costs if adverse effects of supposedly effective interventions introduced into a system have to be straightened out. The predictive validity is acceptable but limited due to methodological issues. Nevertheless, the framework by Grabbe et al. (2022a) provides tools to demonstrate and increase the reliability and validity of an FRAM model, differentiating distinct types of validity for different purposes. The utility of the analysed FRAM model is limited in terms of predictive validity if it is used as an evaluative method. Instead, the utility of the FRAM model as an analytical method is high and invaluable.

It can be concluded that the RPG in terms of FRAM could be bridged and reduced in terms of validity and reliability as well as potential analyst bias in the past and in this thesis due to advancements in guidance and software tools (e.g., FMV, FMI, myFRAM, DynaFRAM), quantification and its related approaches combined with qualitative analysis facilitating a structured analysis and derivation of implications of how to manage variability (e.g., Falegnami et al., 2020; Grabbe et al., 2022b; Hirose & Sawaragi, 2020; Patriarca et al., 2017b), and tools and approaches concerning reliability and validity (e.g., FMI, Grabbe et al., 2022a, see Section 9). Nevertheless, the function identification process and the creation of the FRAM model, as well as the gathering of variability data, is very time- and resource-consuming. One solution to overcome this could be the information technology framework for sharp-end operators' WAD data gathering through a mobile app that Constantino et al. (2020) proposed. Furthermore, the fourth step of FRAM to manage variability can be enhanced by a standardised report delivering basic evaluations in the spirit of FRAM, which would facilitate the communication between decision-makers and the modelers and analysts by telling a reasonable and useful "system story" through the lens of FRAM (Sujan et al., 2023). These structured reports could be created in Microsoft Power BI, enabling, e.g., interactive and dynamic visualisations if doing what-if analyses.

Ultimately, in any safety analysis, a trade-off must be made between the thoroughness of the analysis and the efficiency of completing it. This requires expertise in both the theory or applied method and model and in the application domain, which is pursued by a robust reality-based safety science research, i.e., a science where theory is grounded in rigorous observations of existing practice and practice is based on established theory (Rae et al., 2020). In terms of FRAM, the thoroughness is high despite limited validity and, at the same time, improved efficiency to a satisfying degree, resulting in an appropriate balance. This fact should foster the application of FRAM by practitioners in the industry. However, the practical applicability of FRAM in its ease of use has to be researched and improved further, as Farooqi et al. (2022) claimed. Nevertheless, it is crucial to keep the spirit of FRAM meaningful by preventing it from falling back into the traps of reductionism because systems thinking and complex systems will always require some level of thinking and comprehension by analysts, which requires extensive time and resources.

10.2.4 Research outlook

In future research, the following methodological enhancements are fruitful. First, based on Steckhan et al. (2022), it becomes apparent that a FRAM analysis of road traffic should not only systemically focus on safety issues but also systemically consider multiple functional purposes and demands and their intertwinings. For example, the goal of HAD extends beyond safety to enhance efficiency and comfort (Maurer et al., 2015). Additionally, for AD to be widely accepted, passengers inside the vehicle and individuals interacting with the vehicle externally must trust the automation and embrace the new technology. Unfortunately, these various aspects of system performance are often considered separately, resulting in a fragmented understanding, also known as siloed thinking, where only partial insights are gained (Hollnagel, 2020b). However, these different perspectives are interconnected, necessitating a future synthesis of their analysis based on the concept of Synesis (Hollnagel, 2020b) which involves integrating multiple viewpoints into a comprehensive analysis. A promising approach to implement this in FRAM could be causal loop diagrams (Sterman, 2000) showing reinforcing or balancing effects between different purposes for every function, potentially revealing conflicting goals. Furthermore, multilayer networks (cf. Falegnami et al., 2020) distinguishing different purposes or the Abstraction/Agency framework (Patriarca et al., 2017a) could also be helpful.

Second, estimating the impact of certain conditional factors, including sources of external variability or even internal variability, within the same functional scenario could be valuable (cf. Patriarca et al., 2017b). This could be achieved by determining an influence factor of the respective conditional factor on each function within the model. For example, this could be used to understand the potential impact of weather-related factors such as fog, sudden events such as wildlife traversal, or human factors such as time pressure in the whole system. Thus, one can determine how many functions would be affected and how the *GSV* behaves. Ultimately, the criticality of these conditional factors in the overall system can be considered for different system designs. However, it must be emphasised that this probably changes not only the variability parameterisation of the model but also the model itself, including functions and their couplings.

Third, the FRAM model implicitly includes dynamics but depicts the system behaviour as relatively static, like a snapshot. However, it is required to explicitly capture the dynamic and continuous behaviour, including reciprocity over time which would be more realistic as conceptually supported by Steen et al. (2021). For example, simulating multiple aborted overtakings and their effects would be interesting. DynaFRAM (Salehi et al., 2021a) or the approach by Hirose et al. (2021) using fuzzy reasoning and cellular automaton can investigate such dynamics. In addition, as applied by Patriarca et al. (2017b), Monte-Carlo simulation could be used to show the dynamic effect of different combinations of performance variabilities distinguishing instantiations within one scenario.

Fourth, the new metrics implemented in the semi-quantitative approach have successfully enhanced the calculation and visualisation of interactivity between functions within the system. They have also effectively captured the complex emergence effects of each function. These metrics have served their intended purpose by indicating a function's weight and robustness or tolerance in relation to variability. However, their significance as an influencing parameter, particularly concerning the composition of weighting factors *WaU* and *WaD*, remains a theoretical concept that requires empirical validation in the future. Furthermore, although the calculations currently treat various aspects of couplings equally, except for the propagation factor, it is worth considering a more nuanced approach in the future. This differentiated approach could reveal potential distinct effects resulting from aspects not only qualitatively but also quantitatively.

Lastly, following a many model systems ergonomics approach (Salmon & Read, 2019), it would be reasonable to compare FRAM with other systems ergonomics methods such as Accimap (Svedung & Rasmussen, 2002), cognitive work analysis (CWA, Vicente, 1999), event analysis of systemic teamwork (EAST, Stanton et al., 2013), STAMP (Leveson, 2004), or networked hazard analysis and risk management system (Net-HARMS, Dallat et al., 2018). The comparison should be illustrated through case studies concerning the prospective and holistic safety assessment of automated vehicles. The goal is to create a fruitful toolbox of methods and models that produce diverse but complementary insights to understand the complexity.

10.3 Application in industry

The following describes how FRAM should be integrated into the standardised system engineering V-model (see Figure 10) to develop and prove automated vehicles' safety. FRAM should be mainly used right at the beginning of the earliest stage of concept development as the fundamental basic method providing an analytical tool to derive systems designs or shaping targeted empirical tests. Any model obtained through the FRAM can be used as a basis, either knowledge or integration tool, for other analysis approaches (e.g., FTA, FMEA, HAZOP, STPA). With regard to this, FRAM presents the functional level analysis needed, which, combined with the physical level of analysis, provides the multi-abstraction levels required to overcome the envisioned world problem (Woods & Christoffersen, 2000; Woods & Dekker, 2000). This approach delivers various generic insights for improvement fed back by a cycle to the physical level, which feeds back the consequences (Hirose, 2020). Abstractly, FRAM has the potential to generate bookends where multiple stories can be told (Dekker, 2016; Patriarca et al., 2020) incorporating the learning from all operations (see Figure 21) in order to understand the different ways a system works, i.e., how similar outcomes happen due to different causes and how different outcomes happen due to similar causes. This is required to anticipate what may happen in the future.

More concretely, FRAM can be used to analyse the systemic and holistic functional requirements HAD must fulfill in a specific scenario when entering the road system by identifying the critical functions (i.e., activities) and their interdependencies to produce a desired outcome. This is achieved by functional resonance analysis and its evoking patterns.

Furthermore, FRAM can analytically guide the system design space by assisting in identifying and allocating functions to different agents, i.e., driver or automation, considering emergentisms in order to support a targeted system design. This design space refinement can indirectly reduce the validation effort as the FRAM can provide key insights into the functional dependencies and the impact of variability within the system. This allows for the developing of appropriate and efficient test scenarios that capture potential resonances and evaluate the system's behavior in varying conditions.

Hence, the validation work could be reduced to a reasonable effort by minimising the possible parameter space.

In particular, FRAM should represent the method of choice to analyse STS's complex behaviour, including predominantly scenarios characterised by much interaction between different agents. This predominantly comprises so-called in-traffic and in-vehicle interactions. In this thesis, in-traffic interactions are defined as "situations where the behaviour of at least two road users can be interpreted as being influenced by the possibility that they are both intending to occupy the same region of space at the same time in the near future" (Markkula et al., 2020, p. 737). According to Markkula et al. (2020), interactions in traffic can be distinguished into five distinct prototypical space-sharing conflicts:

- obstructed path
- merging paths
- crossing paths
- constrained and unconstrained head-on paths

Note that when a conflict arises involving more than two road users sharing the same space, the situation may involve multiple prototypes. The following scenarios represent typical examples: mergings between two vehicles as cut-in or cut-out, crossings between vulnerable road users at crosswalks or intersections, crossings between other motorised vehicles at intersections, overtaking, and bottlenecks. It should be emphasised that these scenarios have to be analysed in two ways, i.e., the HAV has an active (inducing) and a passive (involved) role. Instead, in-vehicle interaction contains the driver's and automation's collaboration as shared and traded control to operate the vehicle. Examples are ADASs as LoDA 1 or 2 systems, LoDA 3 in case of takeovers and transitions between different LoDAs in general, or even the level of haptic authority (Abbink et al., 2012) in case of requests to intervene (cf. Inagaki & Sheridan, 2019) exemplifying adaptive automation. Moreover, the H-mode concept (Flemisch et al., 2014) or maneuver-based driver-vehicle cooperation (Franz et al., 2012; Kauer et al., 2010; Walch et al., 2016; 2019), following the LoA and ToA, would also be worth it to be investigated by FRAM. Further, FRAM would be useful to analyse scenarios, including remote assistance or remote driving, helping to identify potential conflicts in the flow of interactions in such complex systems.

Moreover, FRAM and the concept of safety-II could even be used directly for validation following the approach in Section 6. The problem has to be thought of from a different perspective. The current approach to safety approval is based on track distance that focuses on severity, such as accidents. Unfortunately, these events are very unfrequent. This results in the necessity to drive an unfeasible amount of test kilometres – the approval trap. As already discussed comprehensively in Section 2.5, all the common approaches to overcome this issue, e.g., developed in large German or European research projects like PEGASUS (German Aerospace Center [DLR]), VVM (VVM consortium), and L3Pilot (L3Pilot consortium), are still rooted in reductionism focusing on severity, not solving the actual problem, shifting the problem from reality to virtuality (i.e., simulation). This may solve the problem of test kilometres, but a new validity problem occurs. However, if we focus on frequency, we have a proper solution to reduce the test kilometres significantly but still have tests in reality. We need a description of the daily activity of driving and its expected variability, which means one generic case instead of many specific ones that can be used for generalisation.

The data for performance variability of the driver can be obtained one-time, similar to the 100-car study (Dingus et al., 2006), through large-scale data campaigns using monitoring systems in the vehicle such as flight data recorders and tracking technologies of the infrastructure to capture the performance mainly of action functions and simulator studies and interviews to capture the performance of perception and cognitive functions. This data then has to be studied systemically and holistically by FRAM, analysing how the variabilities can propagate through the system using resilience indicators or metrics of the entire system performance. One generic model covers several instantiations. The same can be done for automation. The *GSV* has the potential to provide a system-wide performance metric, i.e., adaptive capacity, making it ultimately possible to compare the driver and automation by one metric, as claimed by Winner (2016). This metric not only compares the driving performance of the driver or the automation in isolation to the rest of the system but also provides a holistic system's driving performance considering the interactions in the entire system.

10.4 Conclusion

The research in this thesis demonstrates that RE, including FRAM and safety-II, is an invaluable approach to assessing the safety of automated driving in road traffic, mainly by studying interactions in view of emergentisms, which are currently neglected by existing approaches but inevitable. In particular, FRAM offers a white-box modelling to facilitate the understanding of how a system usually works in order to comprehend why it can rarely fail. This knowledge can be used to implement targeted interventions to amplify what drivers currently do well in coping with complexity. FRAM argues for more analytical research as a starting point rather than innumerable, indiscriminate empirical testing. However, the current trend of increasing utilisation of AI leads to solutionism or solutioneering (Morozov, 2013; Hollnagel, 2021) which heavily relies on black-box models used for safety-critical safety management resembling a dangerous blind flight by trial-and-error. In general, a FRAM model gives insights where to look in a system by better defining what the problem actually is and to ask better questions but it does not provide explicit solutions per se.

FRAM offers various opportunities and unique benefits that should be used complementary to existing approaches to tackle the multidimensional road safety problem. An one-size-fits-all solution does not exist. Instead, we need an enriched toolbox synthesising different perspectives and approaches. According to Nemeth (2013): “rather than a destination, FRAM is the most recent step […] in understanding complex socio-technical systems”. Thus, FRAM is a promising approach addressing the nitty-gritty - to identify the patterns that facilitate the system’s adaptive capacity inevitable for safe and efficient performance in STSs (Hollnagel, 2016). Therefore, FRAM constitutes an essential missing piece of the puzzle for managing some of the current and future challenges in assessing the safety of automated driving in the complex, dynamic, socio-technical road system.

The primary objective should be enhancing the safety of the entire system by promoting efficient interaction among drivers, machines, and road users. Instead of focusing on ensuring the safety of vehicle automation and providing proof of its safety, the key question becomes how we can utilise automation to design a traffic system that optimises several conflicting goals such as safety, efficiency, and comfort. This research shows possible solutions using FRAM and the perspective of RE in order to tackle the key question. However, due to its high complexity, this work cannot provide complete solutions to the upcoming challenges of introducing automated vehicles in public road transport. Nevertheless, a solid foundation addressing the right problem is laid where enhancements and extensions have to be researched in the future of how to deploy the approach in the industry on a large-scale.

To conclude, a system analysis and design process supported by FRAM with a systems thinking mindset inherently cannot generate final solutions or end-states due to constant changes in the system. Rather it can provide continuous and iterative interventions in the form of different leverage points within the system which may have positive or negative effects. These interventions strive to increase the system's adaptive capacity, i.e., the capability to cope with complexity in normal and abnormal operating conditions. We never fully know the consequences over time and space in the whole system; it depends on the perspective and constraints one might take:

"Once upon a time, there was a Chinese farmer who lost a horse. It ran away. And all the neighbors came around that evening and said, "that's too bad."

And he said, "maybe."

The next day, the horse came back and brought seven wild horses with it. And all the neighbors came around and said, "why that's great, isn't it?"

And he said, "maybe."

The next day his son attempted to tame one of these horses, and was riding it, and was thrown and broke his leg. And all the neighbors came around in the evening and they said, "well, that's too bad, isn't it?"

And the farmer said, "maybe."

The next day conscription officers came around looking for people to join the army and they rejected his son because he had a broken leg. And all the neighbors came around that evening and they said, "well, isn't that wonderful?"

And the farmer said, "maybe.""

– Alan Watts –

References

Abbink, D. A., Carlson, T., Mulder, M., De Winter, J. C., Aminravan, F., Gibo, T. L., & Boer, E. R. (2018). A topology of shared control systems—finding common ground in diversity. IEEE Transactions on Human-Machine Systems, 48(5), 509-525.

Abbink, D. A., Mulder, M., & Boer, E. R. (2012). Haptic shared control: smoothly shifting control authority? Cognition, Technology & Work, 14(1), 19-28.

Accou, B., & Reniers, G. (2019). Developing a method to improve safety management systems based on accident investigations: The SAfety FRactal ANalysis. Safety science, 115, 285-293.

Ackoff, R. L. (1989). From data to wisdom. Journal of applied systems analysis, 16(1), 3-9.

Ackoff, R. L. (1971). Towards a system of systems concepts. Management science, 17(11), 661-671.

Adriaensen, A., Berx, N., Pintelon, L., Costantino, F., Di Gravio, G., & Patriarca, R. (2022). Interdependence Analysis in collaborative robot applications from a joint cognitive functional perspective. International Journal of Industrial Ergonomics, 90, 103320.

Adriaensen, A., Patriarca, R., Smoker, A., & Bergström, J. (2019). A socio-technical analysis of functional properties in a joint cognitive system: a case study in an aircraft cockpit. Ergonomics, 62(12), 1598-1616.

Aeberhard, M., Rauch, S., Bahram, M., Tanzmeister, G., Thomas, J., Pilat, Y., ... & Kaempchen, N. (2015). Experience, results and lessons learned from automated driving on Germany's highways. IEEE Intelligent transportation systems magazine, 7(1), 42-57.

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I., & Shlomo, A. (1977). A Pattern Language: Towns, Buildings, Construction (Vol. 2). Oxford University Press.

Amersbach, C., & Winner, H. (2019). Functional decomposition—A contribution to overcome the parameter space explosion during validation of highly automated driving. Traffic injury prevention, 20(sup1), S52-S57.

Anfara Jr, V. A., Brown, K. M., & Mangione, T. L. (2002). Qualitative analysis on stage: Making the research process more public. Educational researcher, 31(7), 28-38.

Anvarifar, F., Voorendt, M. Z., Zevenbergen, C., & Thissen, W. (2017). An application of the Functional Resonance Analysis Method (FRAM) to risk analysis of multifunctional flood defences in the Netherlands. Reliability Engineering & System Safety, 158, 130-141.

Arenius, M. (2017). Identification of Change Patterns for the Generation of Models of Work-as-Done using Eye-tracking (Vol. 22). Kassel university press GmbH.

Arthur, W. B. (1999). Complexity and the Economy. Science, 284(5411), 107-109.

Asljung, D., Nilsson, J., & Fredriksson, J. (2017). Using extreme value theory for vehicle level safety validation and implications for autonomous vehicles. IEEE Transactions on Intelligent Vehicles, 2(4), 288-297.

(At)City consortium. @CITY. Automated driving in the city. Retrieved from https://www.atcity-online.de/en/ (accessed on 15 October 2022).

Awal Street Journal. Systems Thinking Speech by Dr. Russell Ackoff [Video]. Retrieved from https://www.youtube.com/watch?v=EbLh7rZ3rhU (accessed on 22 August 2021).

Bainbridge, L. (1983). Ironies of automation. In Analysis, design and evaluation of man–machine systems (pp. 129-135). Pergamon.

Banerjee, S. S., Jha, S., Cyriac, J., Kalbarczyk, Z. T., & Iyer, R. K. (2018). Hands off the wheel in autonomous vehicles?: A systems perspective on over a million miles of field data. In *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)* (pp. 586-597). IEEE.

Bansal, P., & Kockelman, K. M. (2017). Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies. Transportation Research Part A: Policy and Practice, 95, 49-63.

Begum, S. (2013). Intelligent driver monitoring systems based on physiological sensor signals: A review. In 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013) (pp. 282-289). IEEE.

Bengler, K., Drüke, J., Hoffmann, S., Manstetten, D., & Neukum, A. (2018). UR:BAN human factors in traffic. *Approaches for Safe, Efficient and Stress-Free Urban Traffic; Springer: Wiesbaden, Germany*.

Bengler, K., Winner, H., & Wachenfeld, W. (2017). No Human–No Cry?. *at-Automatisierungstechnik*, *65*(7), 471-476.

Billings, C. (1993). Aviation Automation. New Jersey: Lawrence Erlbaum.

Björnsdóttir, S. H., Jensson, P., de Boer, R. J., & Thorsteinsson, S. E. (2022). The Importance of Risk Management: What is Missing in ISO Standards?. Risk Analysis, 42(4), 659-691.

Bogdoll, D., Orf, S., Töttel, L., & Zöllner, J. M. (2022). Taxonomy and survey on remote human input systems for driving automation systems. In Advances in Information and Communication: Proceedings of the 2022 Future of Information and Communication Conference (FICC), Volume 2 (pp. 94-108). Cham: Springer International Publishing.

Bonnefon, J. F., Černy, D., Danaher, J., Devillier, N., Johansson, V., Kovacikova, T., ... & Zawieska, K. (2020). Ethics of connected and automated vehicles: Recommendations on road safety, privacy, fairness, explainability and responsibility.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. Science, 352(6293), 1573-1576.

Bonnefon, J. F., Shariff, A., & Rahwan, I. (2019). The trolley, the bull bar, and why engineers should care about the ethics of autonomous cars [point of view]. Proceedings of the IEEE, 107(3), 502-504.

Bose, A., & Ioannou, P. (2001). Evaluation of the environmental effects of intelligent cruise control vehicles. *Transportation research record*, *1774*(1), 90-97.

Bostrom, R. P., & Heinen, J. S. (1977). MIS problems and failures: A socio-technical perspective. Part I: The causes. MIS quarterly, 17-32.

Boudette, N. E. (2017). Tesla's self-driving system cleared in deadly crash. *The New York Times*. Retrieved from https://www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html?searchResultPosition=6 (accessed on 08 March 2021).

Braithwaite, J., Wears, R. L., & Hollnagel, E. (2015). Resilient health care: turning patient safety on its head. International Journal for Quality in Health Care, 27(5), 418-420.

Bubb, H. (2015). Das Regelkreisparadigma der Ergonomie. In Automobilergonomie (pp. 27-65). Springer Vieweg, Wiesbaden.

Campbell, S., O'Mahony, N., Krpalcova, L., Riordan, D., Walsh, J., Murphy, A., & Ryan, C. (2018). Sensor technology in autonomous vehicles: A review. In 2018 29th Irish Signals and Systems Conference (ISSC) (pp. 1-4). IEEE.

Cannas, A., Tedeschi, L. O., Atzori, A. S., & Lunesu, M. F. (2019). How can nutrition models increase the production efficiency of sheep and goat operations?. Animal Frontiers, 9(2), 33-44.

Cao, Y., Griffon, T., Fahrenkrog, F., Schneider, M., Naujoks, F., Tango, F., ... & Shi, E. (2022). L3Pilot-Code of Practice for the Development of Automated Driving Functions.

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.

Carsten, O., Lai, F. C., Barnard, Y., Jamson, A. H., & Merat, N. (2012). Control task substitution in semiautomated driving: Does it matter what aspects are automated? Human factors, 54(5), 747-761.

Checkland, P. (1981). Systems thinking, systems practice, New York: John Wiley & Sons.

Chityala, S., Sobanjo, J. O., Erman Ozguven, E., Sando, T., & Twumasi-Boakye, R. (2020). Driver Behavior at a Freeway Merge to Mixed Traffic of Conventional and Connected Autonomous Vehicles. Transportation Research Record: Journal of the Transportation Research Board, 2674(11), 867–874. https://doi.org/10.1177/0361198120950721.

Christoffersen, K., & Woods, D. D. (2002). How to make automated systems team players. In E. Salas (Ed.), Advances in human performance and cognitive engineering research: Vol. 2. Automation (1sted., pp. 1–12). Amsterdam, Boston: JAI. https://doi.org/10.1016/S1479-3601(02)02003-9.

Cilliers, P. (2002). Why we cannot know complex things completely. Emergence, 4(1-2), 77-84.

Cilliers, P. (1998). Complexity and postmodernism: understanding complex systems. Routledge, London.

Constantino, F., Di Gravio, G., Falegnami, A., Patriarca, R., Tronci, M., De Nicola, A., ... & Villani, M. L. (2020). Crowd sensitive indicators for proactive safety management: A theoretical framework. In Proceedings of the 30th European Safety and Reliability Conference ESREL and 15th Probabilistic Safety Assessment and Management Conference (pp. 1453-1458). Singapore: Research Publishing Services.

Cook, R., Woods, D. D., & Miller, C. (1998). National Health Care Safety Council-A Tale of Two Stories: Contrasting Views of Patient Safety. In A Report from a Workshop on Assembling the Scientific Basis for Progress on Patient Safety. Chicago: National Patient Safety Foundation.

Cooke, D. L., & Rohleder, T. R. (2006). Learning from incidents: from normal accidents to high reliability. System Dynamics Review, 22(3), 213-239.

Cornelissen, M., Salmon, P. M., Stanton, N. A., & McClure, R. (2015). Assessing the 'system' in safe systems-based road designs: using cognitive work analysis to evaluate intersection designs. Accident Analysis & Prevention, 74, 324-338.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. Theory into practice, 39(3), 124-130.

Daimler AG. (2009). Mercedes-Benz präsentiert in Genf Limousine und Coupé der neuen E-Klasse.

Dallat, C., Salmon, P. M., & Goode, N. (2019). Risky systems versus risky people: To what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature. Safety Science, 119, 266-279.

Dallat, C., Salmon, P. M., & Goode, N. (2018). Identifying risks and emergent risks across sociotechnical systems: the NETworked hazard analysis and risk management system (NET-HARMS). Theoretical Issues in Ergonomics Science, 19(4), 456-482.

Das, S., Sun, X., Wang, F., & Leboeuf, C. (2015). Estimating likelihood of future crashes for crash-prone drivers. *Journal of traffic and transportation engineering (English edition)*, *2*(3), 145-157.

De Winter, J. C., & Dodou, D. (2014). Why the Fitts list has persisted throughout the history of function allocation. Cognition, Technology & Work, 16(1), 1-11.

De Winter, J. C., & Hancock, P. A. (2015). Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them? Procedia Manufacturing, 3, 5334-5341.

Dekker, S. (2019). Foundations of safety science: A century of understanding accidents and disasters. Routledge.

Dekker, S. (2016). Patient safety: a human factors approach. CRC Press.

Dekker, S. (2011). Drift into Failure: From Hunting Broken Components to Understanding Complex Systems. Ashgate Publishing, Ltd..

Dekker, S. (2006). Resilience engineering: Chronicling the emergence of confused consensus. In E. Hollnagel, D. D. Woods & N. Leveson (Eds.), Resilience engineering: Concepts and precepts. Hampshire: Ashgate.

Dekker, S., Bergström, J., Amer-Wåhlin, I., & Cilliers, P. (2013). Complicated, complex, and compliant: best practice in obstetrics. Cognition, Technology & Work, 15(2), 189-195.

Dekker, S., Cilliers, P., & Hofmeyr, J. H. (2011). The complexity of failure: Implications of complexity theory for safety investigations. Safety science, 49(6), 939-945.

Dekker, S., Hollnagel, E., Woods, D., & Cook, R. (2008). Resilience Engineering: New directions for measuring and maintaining safety in complex systems. Lund University School of Aviation, 1, 1-6.

Di Fabio, U., Broy, M., Hilgendorf, E., & Nehm, K. (2017). Ethik-Kommission Automatisiertes und Vernetztes Fahren: eingesetzt durch den Bundesminister für Verkehr und digitale Infrastruktur: Bericht. Bundesministerium für Verkehr und digitale Infrastruktur.

Di Maio, P. (2014). Towards a metamodel to support the joint optimization of socio technical systems. Systems, 2(3), 273-296.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J., et al. (2006). The 100-Car Naturalistic Driving Study Phase II – Results of the 100-Car Field Experiment. Report No. DOT HS 810 593. Washington: National Highway Traffic Safety Admin. (NHTSA)

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proceedings of the National Academy of Sciences*, *113*(10), 2636-2641.

Dixit, V. V., Chand, S., & Nair, D. J. (2016). Autonomous vehicles: disengagements, accidents and reaction times. *PLoS one, 11*(12), e0168054.

DMV California. Retrieved from https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing (accessed on 14 Januar 2019).

Dollard, J., Doob, L. W., Miller, N. E., Mowrer, O. H., & Sears, R. R. (1939). *Frustration and aggression*. New Haven: Yale University Press.

Donges, E. (2015). Fahrerverhaltensmodelle. In *Handbuch Fahrerassistenzsysteme* (pp. 17-26). Springer Vieweg, Wiesbaden.

Drösler, J. (1965). Zur Methodik der Verkehrspsychologie. In Psychologie des Straßenwesens; Hoyos, C., Ed.; Huber: Bern, Switzerland; Stuttgart, Germany.

Emery, F.E. (1972). Characteristics of sociotechnical systems. In Davis, L.E., Taylor, J .C. (Eds.), Design for Jobs. Pinguin Books, Hannondsworth. pp. 157-186.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. Human factors, 37(2), 381-394.

Ertrac. (2017). Automated driving roadmap. *ERTRAC Working Group.* Retrieved from https://www.ertrac.org/uploads/documentsearch/id48/ERTRAC_Automated_Driving_2017.pdf. (accessed on 17 April 2020).

Eurocontrol, A. (2022). Unearthing Weak Signals for safer and more efficient socio-technical systems - The Structured Exploration of Complex Adaptations (SECA) method. A white paper.

Eurocontrol, A. (2021a). The Systemic Potentials Management: Building a Basis for Resilient Performance. A white paper.

Eurocontrol, A. (2021b). Patterns in How People Think and Work Importance of Patterns Discovery for Understanding Complex Adaptive Systems. A white paper.

Eurocontrol, A. (2009). White Paper on Resilience Engineering for ATM. Report of the Project Resilience Engineering for ATM.

Fach, M., Baumann, F., Breuer, J., & May, A. (2010). Bewertung der Beherrschbarkeit von Aktiven Sicherheits-und Fahrerassistenzsystemen an den Funktionsgrenzen. *VDI-Berichte*, (2104).

Falegnami, A., Costantino, F., Di Gravio, G., & Patriarca, R. (2020). Unveil key functions in socio-technical systems: mapping FRAM into a multilayer network. Cognition, Technology & Work, 22(4), 877-899.

Farooqi, A., Ryan, B., & Cobb, S. (2022). Using expert perspectives to explore factors affecting choice of methods in safety analysis. Safety science, 146, 105571.

Fastenmeier, W. (2015). Fahrerassistenzsysteme (FAS) und Automatisierung im Fahrzeug–wird daraus eine Erfolgsgeschichte?. Zeitschrift für Verkehrssicherheit, 61(1), 15-25.

Fastenmeier, W., & Gstalter, H. (2007). Driving task analysis as a tool in traffic safety research and practice. Safety Science, 45(9), 952-979.

Favarò, F. M., Eurich, S. O., & Nader, N. (2018, January). Analysis of Disengagements in Autonomous Vehicle Technology. In *2018 Annual Reliability and Maintainability Symposium (RAMS)* (pp. 1-7). IEEE.

Favarò, F. M., Nader, N., Eurich, S. O., Tripp, M., & Varadaraju, N. (2017). Examining accident reports involving autonomous vehicles in California. *PLoS one, 12*(9).

Feldhütter, A. (2021). Effect of Fatigue on Take-Over Performance in Conditionally Automated Driving (Doctoral dissertation, Technische Universität München).

Feldhütter, A., Gold, C., Hüger, A., & Bengler, K. (2016). Trust in automation as a matter of media influence and experience of automated vehicles. In *Proceedings of the Human Factors and Ergonomics Society (HFES) 60th Annual Meeting 2016* (Vol. 60, pp. 2024–2028). https://doi.org/10.1177/1541931213601460

Felkai, R., & Beiderwieden, A. (2011). Schaffen allgemeiner Voraussetzungen der Projektabwicklung. In *Projektmanagement für technische Projekte* (pp. 4-44). Vieweg+ Teubner.

Feltovich, P. J., Hoffman, R. R., Woods, D., & Roesler, A. (2004). Keeping it too simple: How the reductive tendency affects cognitive engineering. IEEE Intelligent Systems, 19(3), 90-94.

Fenton, R. E. (1970). Automatic vehicle guidance and control—A state of the art survey. *IEEE Transactions on Vehicular Technology, 19*(1), 153-161.

Ferreira, P. N., & Cañas, J. J. (2019). Assessing operational impacts of automation using functional resonance analysis method. Cognition, Technology & Work, 21, 535-552.

Fitts, P. M. (1951). Human engineering for an effective air-navigation and traffic-control system.

Flemisch, F., Abbink, D. A., Itoh, M., Pacaux-Lemoine, M. P., & Weßel, G. (2019). Joining the blunt and the pointy end of the spear: towards a common framework of joint action, human–machine cooperation, cooperative guidance and control, shared, traded and supervisory control. Cognition, Technology & Work, 21(4), 555-568.

Flemisch, F. O., Bengler, K., Bubb, H., Winner, H., & Bruder, R. (2014). Towards cooperative guidance and control of highly automated vehicles: H-Mode and Conduct-by-Wire. Ergonomics, 57(3), 343-360.

Flood, M. D., Lemieux, V. L., Varga, M., & Wong, B. W. (2016). The application of visual analytics to financial stability monitoring. Journal of financial stability, 27, 180-197.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. Oxford review, 5.

Franz, B., Kauer, M., Bruder, R., & Geyer, S. (2012). pieDrive-a New Driver-Vehicle Interaction Concept for Maneuver-Based Driving.

Friedrich, B. (2015). Verkehrliche wirkung autonomer fahrzeuge. In *Autonomes Fahren* (pp. 331-350). Springer Vieweg, Berlin, Heidelberg.

Fuller, R. (2005). Towards a general theory of driver behaviour. Accident analysis & prevention, 37(3), 461-472.

Gasser, T. M., Arzt, C., Ayoubi, M., Bartels, A., Bürkle, L., Eier, J., ... & Vogt, W. (2012). Rechtsfolgen zunehmender fahrzeugautomatisierung. Berichte der Bundesanstalt für Straßenwesen. Unterreihe Fahrzeugtechnik, (83).

Gasser, T. M., Schmidt, E. A., Bengler, K., Chiellino, U., Diederichs, F., Eckstein, L., ... & Zeeb, E. (2015). Report on the Need for Research. Round Table on Automated Driving, Federal Ministry of Transport and Digital Infrastructure, BMVI2015.

Geisslinger, M., Poszler, F., Betz, J., Lütge, C., & Lienkamp, M. (2021). Autonomous driving ethics: From trolley problem to ethics of risk. Philosophy & Technology, 34(4), 1033-1055.

Geisslinger, M., Poszler, F. & Lienkamp, M. (2023). An ethical trajectory planning algorithm for autonomous vehicles. *Nature Machine Intelligence*. https://doi.org/10.1038/s42256-022-00607-z

German Aerospace Center. interACT: Designing cooperative interaction of automated vehicles with other road users in mixed environments. Retrieved from https://www.interact-roadautomation.eu/. (accessed on 05 December 2022).

German Aerospace Center. Pegasus Research Project: Securing Automated Driving Effectively. Retrieved from https://www.pegasusprojekt.de/en/about-PEGASUS. (accessed on 05 December 2022).

Gibson, J. J., & Crooks, L. E. (1938). A theoretical field-analysis of automobile-driving. The American journal of psychology, 51(3), 453-471.

Gold, C., Körber, M., Lechner, D., & Bengler, K. (2016). Taking over control from highly automated vehicles in complex traffic situations: The role of traffic density. *Human Factors: the Journal of the Human Factors and Ergonomics Society*, *58*(4), 642–652. https://doi.org/10.1177/0018720816634226

Goldratt, E. M., & Goldratt-Ashlag, E. (2008). The choice. Great Barrington, MA: North River Press.

Golias, J., Yannis, G., & Antoniou, C. (2002). Classification of driver-assistance systems according to their impact on road safety and traffic efficiency. *Transport reviews*, *22*(2), 179-196.

Grabbe, N., Arifagic, A., & Bengler, K. (2022a). Assessing the reliability and validity of an FRAM model: the case of driving in an overtaking scenario. Cognition, Technology & Work, 24(3), 483-508. https://doi.org/10.1007/s10111-022-00701-7

Grabbe, N., Gales, A., Höcher, M., & Bengler, K. (2022b). Functional Resonance Analysis in an Overtaking Situation in Road Traffic: Comparing the Performance Variability Mechanisms between Human and Automation. Safety, 8(1), 3. https://doi.org/10.3390/safety8010003.

Grabbe, N., Höcher, M., Thanos, A., & Bengler , K. (2020a). Safety Enhancement by Automated Driving: What are the Relevant Scenarios. Annual Meeting of the Human Factors and Ergonomics Society 2020.

Grabbe, N., Kellnberger, A., Aydin, B., & Bengler, K. (2020b). Safety of automated driving: the need for a systems approach and application of the Functional Resonance Analysis Method. Safety science, 126, 104665.

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

Griggs, T. & Wakabayashi, D. (2018). How a Self-Driving Uber Killed a Pedestrian in Arizona. *The New York Times*. Retrieved from https://www.nytimes.com/interactive/2018/03/20/us/self-driving-uber-pedestrian-killed.html?searchResultPosition=4. (accessed on 23 July 2021).

Gründl, M. (2005). Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial und die Gestaltung von Fahrerassistenzsystemen (Doctoral dissertation).

Grunwald, A. (2016). Societal risk constellations for autonomous driving. Analysis, historical context and assessment. In Autonomous driving (pp. 641-663). Springer, Berlin, Heidelberg.

Götze, S. (2021). Mercedes-Benz erhält erste Genehmigung für autonome Fahrfunktion. *Spiegel Mobiltät*. Retrieved from https://www.spiegel.de/auto/mercedes-benz-erhaelt-erste-genehmigung-fuer-autonome-fahrfunktion-a-773d17bf-3f3a-4757-9ff8-1bd6a5416721. (accessed on 28 September 2022).

Hale, A. R., & Hovden, J. (1998). Management and culture: the third age of safety. A review of approaches to organizational aspects of safety, health and environment. Occupational injury, 145-182.

Hancock, P. A. (2019). Some pitfalls in the promises of automated and autonomous vehicles. Ergonomics, 62(4), 479-495.

Harvey, C., & Stanton, N. A. (2014). Safety in System-of-Systems: Ten key challenges. Safety science, 70, 358-366.

Hecht, T., Feldhütter, A., Draeger, K., & Bengler, K. (2019). What do you do? An analysis of non-driving related activities during a 60 minutes conditionally automated highway drive. In *International Conference on Human Interaction and Emerging Technologies* (pp. 28-34). Springer, Cham.

Hegeman, G., Brookhuis, K., & Hoogendoorn, S. (2005). Opportunities of advanced driver assistance systems towards overtaking. European Journal of Transport and Infrastructure Research, 5(4).

Heinrich, H. W. (1941). Industrial Accident Prevention. A Scientific Approach. Industrial Accident Prevention. A Scientific Approach., (Second Edition).

Henderson, R. L., & Edwards, M. L. (1987). Driver performance data book. US Department of Transportation, National Highway Traffic Safety Administration.

Hendricks, D.L.; Fell, J.C.; Freedman, M. (2001). *The Relative Frequency of Unsafe Driving Acts in Serious Injury Accidents*; Final report submitted to NHTSA under contract No. DOT NH 22 94 C 05020; Veridian engineering; Springer: Buffalo, NY, USA.

Heylighen, F. (1989). Causality as distinction conservation. A theory of predictability, reversibility, and time order. Cybernetics and Systems: An International Journal, 20(5), 361-384.

Heylighen, F., Cilliers, P., & Gershenson, C. (2007). Complexity and philosophy. In: Bogg J, Geyer R (eds) Complexity, science and society. Radcliffe Publishing, Oxford, pp. 117–13.

Hill, R.; Boult, M.; Sujan, M.; Hollnagel, E.; Slater, D. Predictive Analysis of Complex Systems' Behaviour. 2020. Retrieved from https://www.researchgate.net/profile/David-Slater/publication/343944100_PREDICTIVE_ANALYSIS_OF_COMPLEX_SYSTEMS\T1\textquoteright_BEHAVIOUR_SWIFTFRAM/links/5f4907e0299bf13c5047f8d3/PREDICTIVE-ANALYSIS-OFCOMPLEX-SYSTEMS-BEHAVIOUR-SWIFTFRAM.pdf (accessed on 16 December 2021).

Hill, R., Hollnagel, E. (2016). Instructions for use of the FRAM model visualiser (FMV). Retrieved from http://functionalresonance.com/onewebmedia/FMV_instructions_0.4.0.pdf (accessed on 10 February 2023).

Hirose, T. (2020). Envisioning Emergent Behaviors of Socio-Technical Systems Based on Functional Resonance Analysis Method.

Hirose, T., & Sawaragi, T. (2020). Extended FRAM model based on cellular automaton to clarify complexity of socio-technical systems and improve their safety. Safety science, 123, 104556.

Hirose, T., Sawaragi, T., Nomoto, H., & Michiura, Y. (2021). Functional safety analysis of SAE conditional driving automation in time-critical situations and proposals for its feasibility. Cognition, Technology & Work, 23, 639-657.

Hoeger, R., Amditis, A., Kunert, M., Hoess, A., Flemisch, F., Krueger, H. P., ... & Pagle, K. (2008). Highly automated vehicles for intelligent transport: HAVEit approach. In *ITS World Congress, NY, USA*.

Holland, J. H. (2014). Complexity: A very short introduction. OUP Oxford.

Hollnagel, E. (2021). The many meanings of AI. In: Digitalisation and human performance. HindSight magazine, 33, 14-16. Retrieved from https://skybrary.aero/sites/default/files/bookshelf/32614.pdf (accessed on 09.06.2023)

Hollnagel, E. (2020a). FRAM Model Interpreter. Retrieved from https://functionalresonance.com/onewebmedia/FMI%20basicPlus%20V3.pdf (accessed on 01 March 2021).

Hollnagel, E. (2020b). *Synesis: The Unification of Productivity, Quality, Safety and Reliability*. Routledge.

Hollnagel, E. (2019a). Advancing resilient performance: From instrumental applications to second-order solutions. In REA Symposium on Resilience Engineering Embracing Resilience.

Hollnagel, E. (2019b). FRAM: Setting the Scene. Retrieved from http://functionalresonance.com/framily-meetings/framily%202019.html (accessed on 11 September 2019).

Hollnagel, E. (2018a). From function allocation to function congruence. In Coping with computers in the cockpit (pp. 29-54). Routledge.

Hollnagel, E. (2018b). The Functional Resonance Analysis Method. A brief Guide on how to use the FRAM. Retrieved from http://functionalresonance.com/onewebmedia/FRAM%20Handbook%202018%20v4.pdf (accessed on 11 July 2019).

Hollnagel, E. (2016). The nitty-gritty of human factors. Human factors and ergonomics in practice: Improving system performance and human well-being in the real world, 45-64.

Hollnagel, E. (2014). Safety-I and Safety-II: The Past and Future of Safety Management. CRC Press.

Hollnagel, E. (2012a). FRAM, the Functional Resonance Analysis Method: Modelling Complex Socio-technical Systems. Ashgate Publishing, Ltd..

Hollnagel, E. (2012b). Coping with complexity: past, present and future. Cognition, Technology & Work, 14(3), 199-205.

Hollnagel, E. (2009a). The ETTO Principle: Efficiency-thoroughness Trade-off: why Things that Go Right Sometimes Go Wrong. Ashgate Publishing, Ltd..

Hollnagel, E. (2009b). The four cornerstones of resilience engineering. In C. P. Nemeth, E. Hollnagel, & S. Dekker (Eds.), Resilience engineering perspectives. Volume 2: Preparation and restoration (pp. 117–134). Aldershot, UK: Ashgate.

Hollnagel, E. (2008). Investigations as an impediment to learning. In Hollnagel, E., Nemeth, C.P., Dekker. S. (Eds): Resilience Engineering Perspectives, Volume 1 - Remaining Sensitive to the Possibility of Failure. Ashgate Publishing, Ltd., pp. 259-268.

Hollnagel, E. (2006a). Resilience: The challenge of the unstable. In E. Hollnagel, D. D Woods, & N. Leveson (Eds.), Resilience engineering: Concepts and precepts (pp. 9–17). Aldersho, UK: Ashgate.

Hollnagel, E. (2006b). A function-centred approach to joint driver-vehicle system design. Cognition, Technology & Work, 8, 169-173.

Hollnagel, E., 2004. Barriers and accident prevention Ashgate. Hampshire.

Hollnagel, E. (2002). Understanding accidents-from root causes to performance variability. In Proceedings of the IEEE 7th conference on human factors and power plants (pp. 1-1). IEEE.

Hollnagel, E. (1998). Cognitive reliability and error analysis method (CREAM). Elsevier.

Hollnagel, E., Hounsgaard, J., & Colligan, L. (2014). FRAM-the Functional Resonance Analysis Method: a handbook for the practical use of the method. Centre for Quality, Region of Southern Denmark.

Hollnagel, E., & Speziali, J. (2008). Study on Developments in Accident Investigation Methods: A Survey of the'State-of-the-Art' (No. SKI-R--08-50). Swedish Nuclear Power Inspectorate.

Hollnagel, E., & Woods, D. D. (2005). Joint cognitive systems: Foundations of cognitive systems engineering. CRC press.

Hollnagel, E., & Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. International journal of man-machine studies, 18(6), 583-600.

Hollnagel, E., Woods, D. D., & Leveson, N. (Eds.). (2006). Resilience engineering: Concepts and precepts. Ashgate Publishing, Ltd..

Hovden, J., Albrechtsen, E., & Herrera, I. A. (2010). Is there a need for new theories, models and approaches to occupational accident prevention? Safety Science, 48(8), 950-956.

Hughes, B. P., Anund, A., & Falkmer, T. (2016). A comprehensive conceptual framework for road safety strategies. *Accident Analysis & Prevention, 90*, 13-28.

Hughes, B. P., Newstead, S., Anund, A., Shu, C. C., & Falkmer, T. (2015). A review of models relevant to road safety. Accident Analysis & Prevention, 74, 250-270.

Inagaki, T. (2010). Traffic systems as joint cognitive systems: issues to be solved for realizing human-technology coagency. Cognition, Technology & Work, 12, 153-162.

Inagaki, T. (2003). Adaptive automation: Sharing and trading of control. Handbook of cognitive task design, 8, 147-169.

Inagaki, T., & Sheridan, T. B. (2019). A critique of the SAE conditional driving automation definition, and analyses of options for improvement. Cognition, technology & work, 21, 569-578.

Inagaki, T., & Stahre, J. (2004). Human supervision and control in engineering and music: similarities, dissimilarities, and their implications. Proceedings of the IEEE, 92(4), 589-600.

ISO Standard 26262 (2018) Road vehicles—functional safety. Retrieved from https://www.iso.org/standard/68385.html. (accessed on 21 November 2022).

ISO Standard 21448 (2022) Road vehicles — Safety of the intended functionality. Retrieved from https://www.iso.org/standard/77490.html. (accessed on 21 December 2022).

Jensen, A., & Aven, T. (2018). A new definition of complexity in a risk analysis setting. Reliability Engineering & System Safety, 171, 169-173.

Joint, M. (1995). Road rage. Transport Research Laboratory, TRID Database.

Junietz, P. M. (2019). Microscopic and macroscopic risk metrics for the safety validation of automated driving. (Doctoral dissertation, Technical University of Darmstadt).

Kauer, M., Schreiber, M., & Bruder, R. (2010). How to conduct a car? A design example for maneuver based driver-vehicle interaction. In 2010 IEEE Intelligent Vehicles Symposium (pp. 1214-1221). IEEE.

Khondaker, B., & Kattan, L. (2015). Variable speed limit: A microscopic analysis in a connected vehicle environment. Transportation Research Part C: Emerging Technologies, 58, 146–159.

Kienle, M., Damböck, D., Kelsch, J., Flemisch, F., & Bengler, K. (2009). Towards an H-Mode for highly automated vehicles: Driving with side sticks. In Proceedings of the 1st international conference on automotive user interfaces and interactive vehicular applications (pp. 19-23).

Kim, D. H. (1999). Introduction to systems thinking (Vol. 16). Waltham, MA: Pegasus Communications.

Kirwan, B., & Ainsworth, L. K. (Eds.). (1992). A guide to task analysis: the task analysis working group. CRC press.

Kitajima, S., Shimono, K., Tajima, J., Antona-Makoshi, J., & Uchida, N. (2019). Multi-agent traffic simulations to estimate the impact of automated technologies on safety. Traffic injury prevention, 20(sup1), S58-S64.

Klauer, C., Dingus, T. A., Neale, V. L., Sudweeks, J. D., & Ramsey, D. J. (2006). The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data.

Koopman, P. (2022). Key Ideas: UL 4600 Safety Standard for Autonomous Vehicles.

Kuzminski, P., Eisele, J. S., Garber, N., Schwing, R., Haimes, Y. Y., Li, D., & Chowdhury, M. (1995). Improvement of Highway Safety I: Identification of Causal Factors Through Fault-Tree Modeling 1. Risk analysis, 15(3), 293-312.

Kyriakidis, M., de Winter, J. C., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., ... & Reed, N. (2019). A human factors perspective on automated driving. *Theoretical Issues in Ergonomics Science, 20*(3), 223-249.

Langenberg, J., Bartels, A., & Etemand, A. (2014). Eu-Projekt "AdaptIVe". Ansätze für hochautomatisches Fahren. In VDI-Berichte: Vol. 2223, Fahrerassistenz und Integrierte Sicherheit: 30. Vdi/vw-Gemeinschaftstagung ; Wolfsburg, 14. Und 15. Oktober 2014. Düsseldorf: VDI-Verl.

Larsson, P., Dekker, S. W., & Tingvall, C. (2010). The need for a systems theory approach to road safety. Safety science, 48(9), 1167-1174.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80.

Leveson, N. (2002). System safety engineering: back to the future. Retrieved from: http://sunnyday.mit.edu/book2.pdf. (accessed on 12 May 2023).

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

Leveson, N. (2011). Engineering a Safer World: Systems Thinking Applied to Safety. MIT Press, Cambridge MA.

Leveson, N. G., & Thomas, J. P. (2018). STPA handbook. Cambridge, MA, USA.

Li, W., He, M., Sun, Y., & Cao, Q. (2019). A proactive operational risk identification and analysis framework based on the integration of ACAT and FRAM. Reliability Engineering & System Safety, 186, 101-109.

Li, Y., & Ibanez-Guzman, J. (2020). Lidar for autonomous driving: The principles, challenges, and trends for automotive lidar and perception systems. IEEE Signal Processing Magazine, 37(4), 50-61.

Lintern, G. (2020). Jens Rasmussen's risk management framework. Theoretical Issues in Ergonomics Science, 21(1), 56-88.

Liu, P., Du, Y., Wang, L., & Da Young, J. (2020). Ready to bully automated vehicles on public roads? Accident Analysis & Prevention, 137, 105457. https://doi.org/10.1016/j.aap.2020.105457

Liu, P., Yang, R., & Xu, Z. (2019). How safe is safe enough for self-driving vehicles?. Risk analysis, 39(2), 315-325.

Lundberg, J., Rollenhagen, C., & Hollnagel, E. (2009). What-You-Look-For-Is-What-You-Find–The consequences of underlying accident models in eight accident investigation manuals. Safety science, 47(10), 1297-1311.

L3Pilot consortium. L3Pilot. Piloting Automated Driving on European Roads. Retrieved from https://www.l3pilot.eu/. (accessed on 08 May 2022).

Ma, Z., & Zhang, Y. (2022). Driver-Automated Vehicle Interaction in Mixed Traffic: Types of Interaction and Drivers' Driving Styles. Human factors, 00187208221088358.

MacKinnon, R. J., Pukk-Härenstam, K., Kennedy, C., Hollnagel, E., & Slater, D. (2021). A novel approach to explore Safety-I and Safety-II perspectives in in situ simulations—the structured what if functional resonance analysis methodology. Advances in Simulation, 6(1), 21.

Maier, F. (2013). Wirkpotentiale moderner Assistenzsysteme und Aspekte ihrer Relevanz für die Fahrausbildung (Doctoral dissertation, Institute of Ergonomics, Technische Universität München).

Markkula, G., Madigan, R., Nathanael, D., Portouli, E., Lee, Y. M., Dietrich, A., ... & Merat, N. (2020). Defining interactions: A conceptual framework for understanding interactive behaviour in human and automated road traffic. Theoretical Issues in Ergonomics Science, 21(6), 728-752.

Martens, M. H., & van den Beukel, A. P. (2013). The road to automated driving: Dual mode and human factors considerations. In 16th international IEEE conference on intelligent transportation systems (ITSC 2013) (pp. 2262-2267). IEEE.

Marti, E., De Miguel, M. A., Garcia, F., & Perez, J. (2019). A review of sensor technologies for perception in automated driving. IEEE Intelligent Transportation Systems Magazine, 11(4), 94-108.

Matthaei, R., Reschka, A., Rieken, J., Dierkes, F., Ulbrich, S., Winkle, T., & Maurer, M. (2015). Autonomes Fahren. In *Handbuch Fahrerassistenzsysteme* (pp. 1139-1165). Springer Vieweg, Wiesbaden.

Matthews, G., Sparkes, T. J., & Bygrave, H. M. (1996). Attentional overload, stress, and simulate driving performance. *Human Performance*, *9*(1), 77-101.

Maurer, M. (2000). Flexible Automatisierung von Straßenfahrzeugen mit Rechnersehen. Nummer 443 in *Verkehrstechnik/Fahrzeugtechnik Reihe 12*. VDI–Verlag, Düsseldorf.

Maurer, M., Gerdes, J. C., Lenz, B., & Winner, H. (2015). *Autonomes Fahren: Technische, rechtliche und gesellschaftliche Aspekte* [Autonomous driving: Technical, legal, and societal aspects]. Berlin: Springer Vieweg

Meadows, D. H. (2008). Thinking in systems. London: Earthscan.

Meadows, D. H. (1999). Leverage points: Places to intervene in a system.

Menzel, T., Bagschik, G., & Maurer, M. (2018, June). Scenarios for development, test and validation of automated vehicles. In 2018 IEEE Intelligent Vehicles Symposium (IV) (pp. 1821-1827). IEEE.

Miller, J. H., & Page, S. E. (2007). Complex Adaptive Systems: An Introduction to Computational Models of Social Life.

Miller, D. P., & Swain, A. D. (1987). Human error and human reliability. In G. Salvendy (Ed.), Handbook of human factors (pp. 219–249). New York: Wiley

Morozov, E. (2013). *To save everything, click here: The folly of technological solutionism*. New York: PublicAffairs.

Müller, M. (2018). Automatisches oder autonomes Fliegen? Ein Statusbericht aus der Verkehrsluftfahrt. In *Workshop Fahrerassistenzsysteme und automatisiertes Fahren.* Uni-DAS e.V. Walting.

Näätänen, R. & Summala, H. (1976). A model for the role of motivational factors in drivers' decision making. Accident Analysis & Prevention, 6 (3-4), pp. 243-261.

Nemeth, C. (2013). Erik Hollnagel: FRAM: The functional resonance analysis method, modeling complex socio-technical systems. Cogn. Technol. Work 15, 117–118. https://doi.org/10.1007/s10111-012-0246-3.

Nemeth, C., Wears, R., Woods, D., Hollnagel, E., & Cook, R. (2008). Minding the Gaps: Creating Resilience in Health Care. Advances in Patient Safety: New Directions and Alternative Approaches, pp. 1-13. doi:NBK43670

Newnam, S., & Goode, N. (2015). Do not blame the driver: A systems analysis of the causes of road freight crashes. Accident Analysis & Prevention, 76, 141-151.

Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, Ba and leadership: a unified model of dynamic knowledge creation. Long range planning, 33(1), 5-34.

Noy, I. Y., Shinar, D., & Horrey, W. J. (2018). Automated driving: Safety blind spots. Safety science, 102, 68-78.

OECD, 1990. Behavioural adaptations to changes in the road transport system. Organization for Economic Co-operation and Development, Road Transport Research, Paris.

Ombredane, A., & Faverge, J. M. (1955). L'analyse du travail. Paris: Presses Universitaires de France.

Otte, D., Pund, B., & Jänsch, M. (2009). A new approach of accident causation analysis by seven steps ACASS. In Proceedings of the International Technical Conference on the Enhanced Safety of Vehicles, Stuttgart, Germany; National Highway Traffic Safety Administration: Washington, DC, USA, Volume 2009.

Ottino, J. M. (2003). Complex systems. American Institute of Chemical Engineers. AIChE Journal, 49(2), 292.

Pacaux-Lemoine, M. P., & Flemisch, F. (2019). Layers of shared and cooperative control, assistance, and automation. Cognition, Technology & Work, 21, 579-591.

Page, S. E. (2008). Uncertainty, difficulty, and complexity. Journal of Theoretical Politics, 20(2), 115-149.

Papadimitriou, E., Afghari, A. P., Tselentis, D., & van Gelder, P. (2022). Road-safety-II: Opportunities and barriers for an enhanced road safety vision. Accident Analysis & Prevention, 174, 106723.

Papadimitriou, E., Schneider, C., Tello, J. A., Damen, W., Vrouenraets, M. L., & Ten Broeke, A. (2020). Transport safety and human factors in the era of automation: What can transport modes learn from each other?. *Accident Analysis & Prevention, 144*, 105656.

Papakostopoulos, V., Marmaras, N., & Nathanael, D. (2017). The "field of safe travel" revisited: interpreting driving behaviour performance through a holistic approach. Transport reviews, 37(6), 695-714.

Parasuraman, R. (2011). Neuroergonomics: Brain, cognition, and performance at work. Current directions in psychological science, 20(3), 181-186.

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human factors, 39*(2), 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Parker, D., Lajunen, T., & Stradling, S. (1998). Attitudinal predictors of interpersonally aggressive violations on the road. *Transportation Research Part F: Traffic Psychology and Behaviour*, *1*(1), 11-24.

Pasmore, W., Francis, C., Haldeman, J., & Shani, A. (1982). Sociotechnical systems: A North American reflection on empirical studies of the seventies. Human relations, 35(12), 1179-1204.

Patriarca, R. (2017). Developing Risk and Safety Management Methods for Complex Sociotechnical Systems: From Newtonian Reasoning to Resilience Engineering. (Doctoral dissertation, Sapienza University of Rome).

Patriarca, R., Bergström, J., & Di Gravio, G. (2017a). Defining the functional resonance analysis space: Combining Abstraction Hierarchy and FRAM. Reliability Engineering & System Safety, 165, 34-46.

Patriarca, R., Del Pinto, G., Di Gravio, G., Constantino, F., 2018. FRAM for systemic accident analysis: a matrix representation of functional resonance. Int. J. Reliab. Qual. Saf. Eng. 25 (01), 1850001.

Patriarca, R., Di Gravio, G., & Costantino, F. (2017b). A Monte Carlo evolution of the Functional Resonance Analysis Method (FRAM) to assess performance variability in complex systems. Safety science, 91, 49-60.

Patriarca, R., Di Gravio, G., & Costantino, F. (2017c). myFRAM: An open tool support for the functional resonance analysis method. In 2017 2nd International Conference on System Reliability and Safety (ICSRS) (pp. 439-443). IEEE.

Patriarca, R., Di Gravio, G., Woltjer, R., Costantino, F., Praetorius, G., Ferreira, P., & Hollnagel, E. (2020). Framing the FRAM: A literature review on the functional resonance analysis method. Safety Science, 129, 104827.

Patriarca, R., Falegnami, A., Costantino, F., Di Gravio, G., De Nicola, A., & Villani, M. L. (2021). WAx: An integrated conceptual framework for the analysis of cyber-socio-technical systems. Safety science, 136, 105142.

Patterson, E. S., Woods, D. D., Cook, R. I., & Render, M. L. (2007). Collaborative cross-checking to enhance resilience. Cognition, Technology & Work, 9, 155-162.

Paulweber, M. (2017). Validation of highly automated safe and secure systems. In Automated Driving (pp. 437-450). Springer, Cham.

Pavard, B., & Dugdale, J. (2006). The contribution of complexity theory to the study of socio-technical cooperative systems. In Unifying themes in complex systems (pp. 39-48). Springer, Berlin, Heidelberg.

Pereira AG (2013). Introduction to the Use of FRAM on the effectiveness assessment of a radiopharmaceutical dispatches process. In: International nuclear Atlantic conference.

Perrow, C. (1984). Normal accidents: living with high-risk technologies. New York: Basic Books.

Pidgeon, N. (2010). Systems thinking, culture of reliability and safety. Civil Engineering and Environmental Systems, 27(3), 211-217.

Pomerleau, D., & Jochem, T. (1996). Rapidly adapting machine vision for automated vehicle steering. *IEEE expert, 11*(2), 19-27.

Popiv, D., Rommerskirchen, C., Bengler, K., Duschl, M., & Rakic, M. (2010). Effects of assistance of anticipatory driving on driver's behaviour during deceleration phases. In *European conference on human centered design for intelligent transport systems* (pp. 133-145). HUMANIST Publications Lyon.

Preuk, K., Stemmler, E., Schießl, C., & Jipp, M. (2016). Does assisted driving behavior lead to safety-critical encounters with unequipped vehicles' drivers?. Accident Analysis & Prevention, 95, 149-156.

Provan, D. J., Dekker, S. W., & Rae, A. J. (2017). Bureaucracy, influence and beliefs: A literature review of the factors shaping the role of a safety professional. Safety science, 98, 98-112.

Punke, M., Menzel, S., Werthessen, B., Stache, N., & Höpfl, M. (2016). Automotive camera (hardware). In H. Winner, S. Hakuli, F. Lotz, & C. Singer (Eds.), Handbook of driver assistance systems: Basic information, components and systems for active safety and comfort (pp. 431–460). Cham: Springer Reference.

Qureshi, Z.H. (2007). A review of accident modelling approaches for complex socio-technical systems. In Proceedings of the 1757 twelfth Australian Workshop on Safety Critical Systems and Software and Safety-Related Programmable Systems, Adeleide, Australia, 30–31 August 2007; Australian Computer Society, Inc.: Darlinghurst, Australia; Volume 86, pp. 47–59.

Rae, A. J., & Alexander, R. D. (2017). Probative blindness and false assurance about safety. Safety science, 92, 190-204.

Rae, A., Provan, D., Aboelssaad, H., & Alexander, R. (2020). A manifesto for Reality-based Safety Science. Safety science, 126, 104654.

Rasmussen, J. (1997). Risk management in a dynamic society: a modelling problem. Safety science, 27(2-3), 183-213.

Rasmussen, J. (1983). Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE transactions on systems, man, and cybernetics*, (3), 257-266.

Rasmussen, J. (1979). On the structure of knowledge-a morphology of mental models in a man-machine system context. RISOE NATIONAL LAB ROSKILDE (DENMARK).

Rasmussen, J., & Lind, M. (1981). Coping with Complexity. Risø National Laboratory: Roskilde, Denmark.

Read, G. J., Beanland, V., Lenné, M. G., Stanton, N. A., & Salmon, P. M. (2017). Systems Thinking in Transport Analysis and Design. In Integrating Human Factors Methods and Systems Thinking for Transport Analysis and Design (pp. 3-17). CRC Press.

Reason, J. (1990). Human error. Cambridge university press.

Reichart, G. (2000). *Menschliche Zuverlässigkeit beim Führen von Kraftfahrzeugen–Möglichkeiten der Analyse und Bewertung* (Doctoral dissertation, Dissertation am Lehrstuhl für Ergonomie der TU München).

Roberts, J., Hodgson, R., & Dolan, P. (2011). "It's driving her mad": Gender differences in the effects of commuting on psychological health. *Journal of Health Economics*, *30*(5), 1064–1076. https://doi.org/10.1016/j.jhealeco.2011.07.006

Roesener, C., Harth, M., Weber, H., Josten, J., & Eckstein, L. (2018, November). Modelling human driver performance for safety assessment of road vehicle automation. In 2018 21st International Conference on Intelligent Transportation Systems (ITSC) (pp. 735-741). IEEE.

RWTH Aachen. UNICARagil: DISRUPTIVE MODULAR ARCHITECTURE FOR AGILE AUTOMATED VEHICLE CONCEPTS. Retrieved from https://www.unicaragil.de/en/. (accessed on 19 June 2021).

Sacks, D., Murray, O., & Brody, L. R. (2014). Encyclopedia of the ancient Greek world. Infobase Publishing.

SAE International (2021). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. (J3016).

SAE International (2016). Cybersecurity Guidebook for Cyber-Physical Vehicle Systems. (J3061)

Saffarian, M., De Winter, J. C., & Happee, R. (2012). Automated driving: human-factors issues and design solutions. In Proceedings of the human factors and ergonomics society annual meeting (Vol. 56, No. 1, pp. 2296-2300). Sage CA: Los Angeles, CA: Sage Publications.

Salehi, V., Smith, D., Veitch, B., & Hanson, N. (2021a). A dynamic version of the FRAM for capturing variability in complex operations. MethodsX, 8, 101333.

Salehi, V., Veitch, B., & Smith, D. (2021b). Modeling complex socio-technical systems using the FRAM: A literature review. Human factors and ergonomics in manufacturing & service industries, 31(1), 118-142.

Salmon, P. M., & Lenné, M. G. (2015). Miles away or just around the corner? Systems thinking in road safety research and practice. Accident analysis and prevention, 74, 243-249.

Salmon, P. M., McClure, R., & Stanton, N. A. (2012). Road transport in drift? Applying contemporary systems thinking to road safety. Safety science, 50(9), 1829-1838.

Salmon, P. M., & Read, G. J. (2019). Many model thinking in systems ergonomics: a case study in road safety. Ergonomics, 62(5), 612-628.

Salmon, P. M., Read, G. J., Stanton, N. A., & Lenné, M. G. (2013). The crash at Kerang: Investigating systemic and psychological factors leading to unintentional non-compliance at rail level crossings. Accident Analysis & Prevention, 50, 1278-1288.

Salmon, P. M., Read, G. J., Walker, G. H., Stevens, N. J., Hulme, A., McLean, S., & Stanton, N. A. (2022). Methodological issues in systems Human Factors and Ergonomics: Perspectives on the research–practice gap, reliability and validity, and prediction. Human Factors and Ergonomics in Manufacturing & Service Industries, 32(1), 6-19.

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.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. Human factors, 37(1), 5-19.

Sarter, N.B., Woods, D.D., & Billings, C.E. (1997). Automation surprises. In Handbook of Human Factors and Ergonomics; Wiley: New York, NY, USA; pp. 1926–1943.

Schoettle, B., & Sivak, M. (2015). A preliminary analysis of real-world crashes involving self-driving vehicles. University of Michigan Transportation Research Institute.

Schuldt, F., Saust, F., Lichte, B., Maurer, M., & Scholz, S. (2013). Effiziente systematische Testgenerierung für Fahrerassistenzsysteme in virtuellen Umgebungen. *Automatisierungssysteme, Assistenzsysteme und Eingebettete Systeme Für Transportmittel*.

Schnieder, E., & Schnieder, L. (2013). Verkehrssicherheit. Springer Berlin Heidelberg.

Scott-Parker, B., Goode, N., & Salmon, P. (2015). The driver, the road, the rules… and the rest? A systems-based approach to young driver road safety. Accident Analysis & Prevention, 74, 297-305.

Shahian Jahromi, B., Tulabandhula, T., & Cetin, S. (2019). Real-time hybrid multi-sensor fusion framework for perception in autonomous vehicles. Sensors, 19(20), 4357.

Shalev-Shwartz, S., Shammah, S., & Shashua, A. (2017). On a formal model of safe and scalable self-driving cars. arXiv preprint arXiv:1708.06374.

Shariff, A., Bonnefon, J. F., & Rahwan, I. (2021). How safe is safe enough? Psychological mechanisms underlying extreme safety demands for self-driving cars. Transportation research part C: emerging technologies, 126, 103069.

Shergold, I., Wilson, M., & Parkhurst, G. (2016). *The mobility of older people, and the future role of connected autonomous vehicles* (Project Report). Bristol. Retrieved from http://eprints.uwe.ac.uk/31998. (accessed on 14 October 2022).

Sheridan, T. B. (1991). Automation, authority and angst—revisited. In Proceedings of the Human Factors Society Annual Meeting (Vol. 35, No. 1, pp. 2-6). Sage CA: Los Angeles, CA: SAGE Publications.

Sheridan, T. B. (1992). *Telerobotics, automation, and human supervisory control*. Cambridge, MA: MIT Press.

Shinar, D. (1998). Aggressive driving: the contribution of the drivers and the situation. *Transportation Research Part F: traffic psychology and behaviour*, *1*(2), 137-160.

Shladover, S. E., & Nowakowski, C. (2019). Regulatory challenges for road vehicle automation: Lessons from the california experience. Transportation research part A: policy and practice, 122, 125-133.

Shladover, Steven E., Dongyan Su, and Xiao-Yun Lu. (2012). Impacts of cooperative adaptive cruise control on freeway traffic flow. *Transportation Research Record* 2324.1: 63-70.

Shorrock S. (2016). The varieties of human work. Retrieved from https://humanisticsystems.com/2016/12/05/the-varieties-of-human-work/. (accessed on 12 May 2023).

Shorrock S. (2022). Getting a Handle on Three Zones of Performance. Retrieved from https://humanisticsystems.com/2022/12/30/getting-a-handle-on-three-zones-of-performance/. (accessed on 12 May 2023).

Singh, S. (2015). Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey; (No. DOT HS 812 115); NHTSA’s National Center for Statistics and Analysis: Washington, DC, USA.

Sivak, M., & Schoettle, B. (2015). Road safety with self-driving vehicles: General limitations and road sharing with conventional vehicles. University of Michigan, Ann Arbor, Transportation Research Institute.

Slovick, M. (2021). World's First Level 3 Self-Driving Production Car Now Available in Japan. *Electronic Design*. Retrieved from https://www.electronicdesign.com/markets/automotive/article/21158656/electronic-design-worlds-first-level-3-selfdriving-production-car-now-available-in-japan. (accessed on 16 November 2022).

Skolnik, M. I. (1962). Introduction to radar. Radar handbook, 2, 21.

Skyttner, L. (2005). General systems theory: Problems, perspectives, practice. World scientific.

Skyttner, L. (2001). General systems theory: ideas & applications. World Scientific.

Stanton, N. A., Brown, J. W., Revell, K. M., Kim, J., Richardson, J., Langdon, P., ... & Thompson, S. (2022). OESDs in an on-road study of semi-automated vehicle to human driver handovers. Cognition, Technology & Work, 24(2), 317-332.

Stanton, N. A., & Marsden, P. (1996). From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. *Safety science, 24*(1), 35-49.

Stanton, N. A., Salmon, P. M., Rafferty, L. A., Walker, G. H., Baber, C., & Jenkins, D. P. (2013). Human Factors Methods: A Practical Guide for Engineering and Design. CRC Press.

Stanton, N. A., & Young, M. S. (2003). Giving ergonomics away? The application of ergonomics methods by novices. Applied Ergonomics, 34(5), 479-490.

Staron, M., & Durisic, D. (2017). Autosar standard. In Automotive Software Architectures (pp. 81-116). Springer, Cham.

Starr, C. (1969). Social benefit versus technological risk: what is our society willing to pay for safety?. Science, 165(3899), 1232-1238.

Ständer, T. (2010). *Eine modellbasierte Methode zur Objektivierung der Risikoanalyse nach ISO 26262* (Doctoral dissertation).

Steckhan, L., Spiessl, W., Quetschlich, N., & Bengler, K. (2022). Beyond SAE J3016: New Design Spaces for Human-Centered Driving Automation. In International Conference on Human-Computer Interaction (pp. 416-434). Springer, Cham.

Steen, R., Patriarca, R., & Di Gravio, G. (2021). The chimera of time: Exploring the functional properties of an emergency response room in action. Journal of Contingencies and Crisis Management, 29(4), 399-415.

Sterman, J. D. (2000). Business Dynamics: Systems Thinking and Modeling for a Complex World. Boston: Irwin McGraw-Hill.

Sujan, M., Pickup, L., de Vos, M. S., Patriarca, R., Konwinski, L., Ross, A., & McCulloch, P. (2023). Operationalising FRAM in Healthcare: A critical reflection on practice. Safety Science, 158, 105994.

Summala, H. & Näätänen, R. (1988). The zero-risk theory and overtaking decisions. In: Rothengatter, J.A. & De Bruin, R.A. (Eds.), Road user behaviour, Theory & research (pp. 82-92). Assen/Maastricht: Van Gorcum.

Taylor, C. C. W. (1999). The Atomists, Leucippus and Democritus: Fragments: a Text and Translation with a Commentary (Vol. 36). University of Toronto Press.

Taylor, F. W. (1911). Principles of Scientific Management. New York: Norton.

Tener, F., & Lanir, J. (2022). Driving from a Distance: Challenges and Guidelines for Autonomous Vehicle Teleoperation Interfaces. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

Thomas, J., Sgueglia, J., Suo, D., Leveson, N., Vernacchia, M., & Sundaram, P. (2015). An integrated approach to requirements development and hazard analysis. SAE Technical Paper, (2015-01), 0274.

Tian, J., Wu, J., Yang, Q., & Zhao, T. (2016). FRAMA: A safety assessment approach based on Functional Resonance Analysis Method. Safety science, 85, 41-52.

Tientrakool, P., Ho, Y. C., & Maxemchuk, N. F. (2011, September). Highway capacity benefits from using vehicle-to-vehicle communication and sensors for collision avoidance. In *2011 IEEE Vehicular Technology Conference (VTC Fall)* (pp. 1-5). IEEE.

Trende, A., Unni, A., Weber, L., Rieger, J. W., & Luedtke, A. (2019). An investigation into human-autonomous vs. human-human vehicle interaction in time-critical situations. Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments. https://doi.org/10.1145/3316782.3321544

Trist, E. L., & Bamforth, K. W. (1951). Some social and psychological consequences of the longwall method of coal-getting: An examination of the psychological situation and defences of a work group in relation to the social structure and technological content of the work system. Human relations, 4(1), 3-38.

Tsugawa, S. (1994). Vision-based vehicles in Japan: Machine vision systems and driving control systems. *IEEE Transactions on industrial electronics, 41*(4), 398-405.

Underwood, P. (2013). Examining the systemic accident analysis research-practice gap (Doctoral dissertation, Loughborough University).

Underwood, P., & Waterson, P. (2012). A critical review of the STAMP, FRAM and Accimap systemic accident analysis models. Advances in human aspects of road and rail transportation; pp. 385-394.

UN ECE R131 (2013). Uniform provisions concerning the approval of motor vehicles with regard to the Advanced Emergency Braking Systems (AEBS).

Van Arem, B., Van Driel, C. J., & Visser, R. (2006). The impact of cooperative adaptive cruise control on traffic-flow characteristics. *IEEE Transactions on intelligent transportation systems*, *7*(4), 429-436.

Vaughan, D. (1996). The Challenger launch decision: Risky technology, culture, and deviance at NASA. University of Chicago press.

Velasco-Hernandez, G., Barry, J., & Walsh, J. (2020). Autonomous driving architectures, perception and data fusion: A review. In 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP) (pp. 315-321). IEEE.

Vicente, K. J. (1999). Cognitive work analysis: Toward safe, productive, and healthy computer-based work. CRC press.

Visser, E., Pijl, Y. J., Stolk, R. P., Neeleman, J., & Rosmalen, J. G. (2007). Accident proneness, does it exist? A review and meta-analysis. Accident Analysis & Prevention, 39(3), 556-564.

Von Bertalanffy, L. (1968). General system theory. New York: George Braziller. Inc. THE.

Von Bertalanffy, L. (1950). The theory of open systems in physics and biology. Science, 111(2872), 23-29.

VVM consortium. VVM: Verification Valiation Methods. Retrieved from https://www.vvm-projekt.de/en/. (accessed on 18 December 2022).

Wachenfeld, W. H. K. (2017). How stochastic can help to introduce automated driving. (Doctoral dissertation, TU Darmstadt).

Wachenfeld, W., & Winner, H. (2016). The release of autonomous vehicles. In *Autonomous driving* (pp. 425-449). Springer, Berlin, Heidelberg.

Wagner, P. (2015). Steuerung und Management in einem Verkehrssystem mit autonomen Fahrzeugen. In *Autonomes Fahren* (pp. 313-330). Springer Vieweg, Berlin, Heidelberg.

Walch, M., Sieber, T., Hock, P., Baumann, M., & Weber, M. (2016). Towards cooperative driving: Involving the driver in an autonomous vehicle's decision making. In Proceedings of the 8th international conference on automotive user interfaces and interactive vehicular applications (pp. 261-268).

Walch, M., Woide, M., Mühl, K., Baumann, M., & Weber, M. (2019). Cooperative overtaking: Overcoming automated vehicles' obstructed sensor range via driver help. In Proceedings of the 11th international conference on automotive user interfaces and interactive vehicular applications (pp. 144-155).

Walker, G. H., Stanton, N. A., & Salmon, P. M. (2015). Lessons from aviation. In Walker, G. H., Stanton, N. A., & Salmon, P. M. (Eds.), *Human Factors in Automotive Engineering and Technology* (pp. 27-38). Ashgate Publishing, Ltd..

Walker, G. H., N. A. Stanton, P. M. Salmon, and D. P. Jenkins. (2009). Command and Control: The Sociotechnical Perspective. Aldershot: Ashgate.

Walker, G. H., Stanton, N. A., Salmon, P. M., Jenkins, D. P., & Rafferty, L. (2010). Translating concepts of complexity to the field of ergonomics. Ergonomics, 53(10), 1175-1186.

Wang, C., & Winner, H. (2019). Overcoming challenges of validation automated driving and identification of critical scenarios. In 2019 IEEE Intelligent Transportation Systems Conference (ITSC) (pp. 2639-2644). IEEE.

Watson, H. A. (1961). BT Laboratories,“Launch control safety study,” Bell Telephone Laboratories, Murray Hill. NJ, Tech. Rep.

Watzenig, D., & Horn, M. (2017a). *Automated driving: safer and more efficient future driving*. Springer.

Watzenig, D., & Horn, M. (2017b). Introduction to automated driving. In D. Watzenig & M. Horn (Eds.), *Automated driving: Safer and more efficient future driving* (pp. 3–16). Cham, s.l.: Springer International Publishing.

Waymo LLC. Waymo One. Retrieved from https://waymo.com/waymo-one/. (accessed on 13 January 2023).

Weick, K. E. (2011). Organizing for transient reliability: The production of dynamic non-events. Journal of contingencies and crisis management, 19(1), 21-27.

Weitzel, A., Winner, H., Peng, C., Geyer, S., Lotz, F., & Sefati, M. (2014). Absicherungsstrategien für Fahrerassistenzsysteme mit Umfeldwahrnehmung.

Wendt, Z., & Cook, J. S. (2018). Saved by the Sensor: Vehicle Awareness in the Self-Driving Age. Machine Design. Retrieved from https://www.machinedesign.com/mechanical-motion-systems/article/21836344/saved-by-the-sensor-vehicle-awareness-in-theselfdriving-age (accessed on 20 May 2020).

Wickens, C. D. (1995). Designing for situation awareness and trust in automation. IFAC Proceedings Volumes, 28(23), 365-370.

Wickens, C. D., Gordon, S. E., Liu, Y., & Lee, J. (2003). An introduction to human factors engineering (Vol. 2). Upper Saddle River, NJ: Pearson Prentice Hall.

Wienen, H. C., Bukhsh, F. A., Vriezekolk, E., & Wieringa, R. J. (2017). Accident analysis methods and models-a systematic literature review. Centre Telematics Inf Technol.

Wiener, N. (1965). Cybernetics: or the Control and Communication in the Animal and the Machine, Cambridge, MA: MIT Press.

Wiener, E. L., & Curry, R. E. (1980). Flight-deck automation: Promises and problems. *Ergonomics, 23*(10), 995-1011.

Wilde, G. J. (1982). The theory of risk homeostasis: implications for safety and health. Risk analysis, 2(4), 209-225.

Winner, H., Hakuli, S., Lotz, F., & Singer, C. (2016). Handbook of driver assistance systems. Amsterdam, The Netherlands:: Springer International Publishing.

Winner, H. (2016). Quo vadis, FAS?. In Handbook of Driver Assistance Systems (pp. 1557-1584). Springer Vieweg, Wiesbaden.

Woods, D. D. (2019). Essentials of resilience, revisited. In Handbook on resilience of socio-technical systems, pp. 52-65.

Woods, D. D. (2003). Creating foresight: How resilience engineering can transform NASA’s approach to risky decision making. Work, 4(2), 137-144.

Woods, D. D., & Branlat, M. (2017). Basic patterns in how adaptive systems fail. In Resilience engineering in practice (pp. 127-143). CRC Press.

Woods, D., & Christoffersen, K. (2000). Balancing practice-centered research and design. Cognitive Systems Engineering in Military Aviation Domains: An Introductory Primer, 12.

Woods, D., & Dekker, S. (2000). Anticipating the effects of technological change: a new era of dynamics for human factors. Theoretical issues in ergonomics science, 1(3), 272-282.

Woods, D.D., S.W.A. Dekker, R.I. Cook, L.J. Johannesen, and N.B. Sarter. (2010). Behind Human Error. Aldershot: Ashgate Publishing Co.

World Health Organization. (2020). *Global Status Report on Road Safety*; WHO Library Cataloguing-in-Publication Data; WHO: Geneva, Switzerland.

Wu, C., Zhao, G., & Ou, B. (2011). A fuel economy optimization system with applications in vehicles with human drivers and autonomous vehicles. Transportation Research Part D: Transport and Environment, 16(7), 515–524.

Yeong, D. J., Velasco-Hernandez, G., Barry, J., & Walsh, J. (2021). Sensor and sensor fusion technology in autonomous vehicles: A review. Sensors, 21(6), 2140.

Young, K. L., & Salmon, P. M. (2015). Sharing the responsibility for driver distraction across road transport systems: A systems approach to the management of distracted driving. Accident Analysis & Prevention, 74, 350-359.

ZENTEC Zentrum für Technologie, Existenzgründung und Cooperation GmbH. Projekt Ko-HAF. Retrieved from https://www.ko-haf.de/startseite/. (accessed on 26 November 2022).

Zhang, Y., Angell, L., & Bao, S. (2021). A fallback mechanism or a commander? A discussion about the role and skill needs of future drivers within partially automated vehicles. Transportation research interdisciplinary perspectives, 9, 100337.

Zhang, Y., Lintern, G., Gao, L., & Zhang, Z. (2021). A Study on Functional Safety, SOTIF and RSS from the Perspective of Human-Automation Interaction (No. 2021-01-0858). SAE Technical Paper.

Appendix

A Overview of remaining critical paths

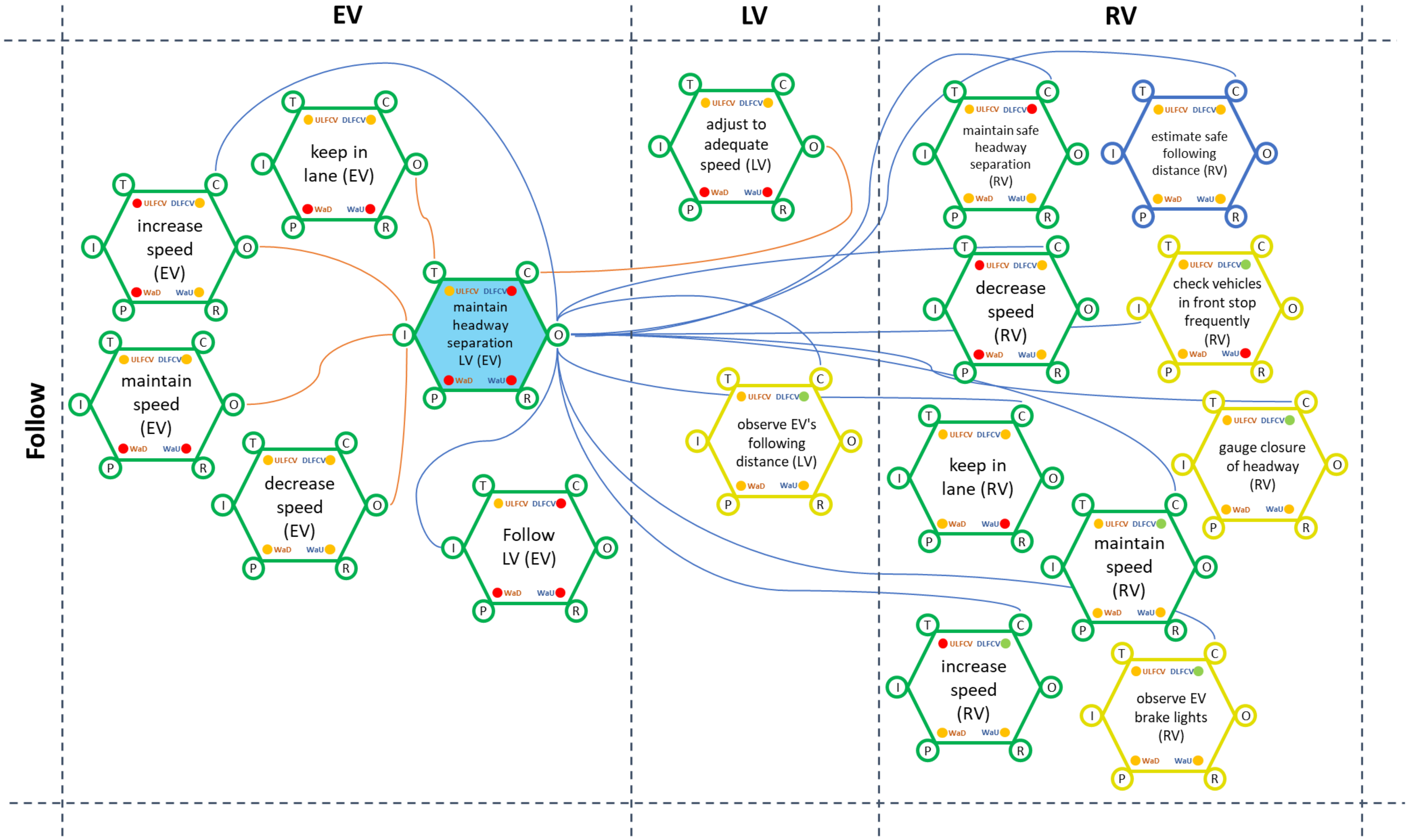

EV
LV
RV
Follow
increase speed (EV)
keep in lane (EV)
maintain speed (EV)
decrease speed (EV)
maintain headway separation LV (EV)
Follow LV (EV)
adjust to adequate speed (LV)
observe EV's following distance (LV)
maintain safe headway separation (RV)
estimate safe following distance (RV)
decrease speed (RV)
check vehicles in front stop frequently (RV)
keep in lane (RV)
gauge closure of headway (RV)
maintain speed (RV)
increase speed (RV)
observe EV brake lights (RV)
T
C
O
R
P
I
ULFCV
DLFCV
WaD
WaU

Figure B1. The critical path of the function “maintain headway separation (EV)” in scenario one. It is the same for scenario three but with different values of the metrics.

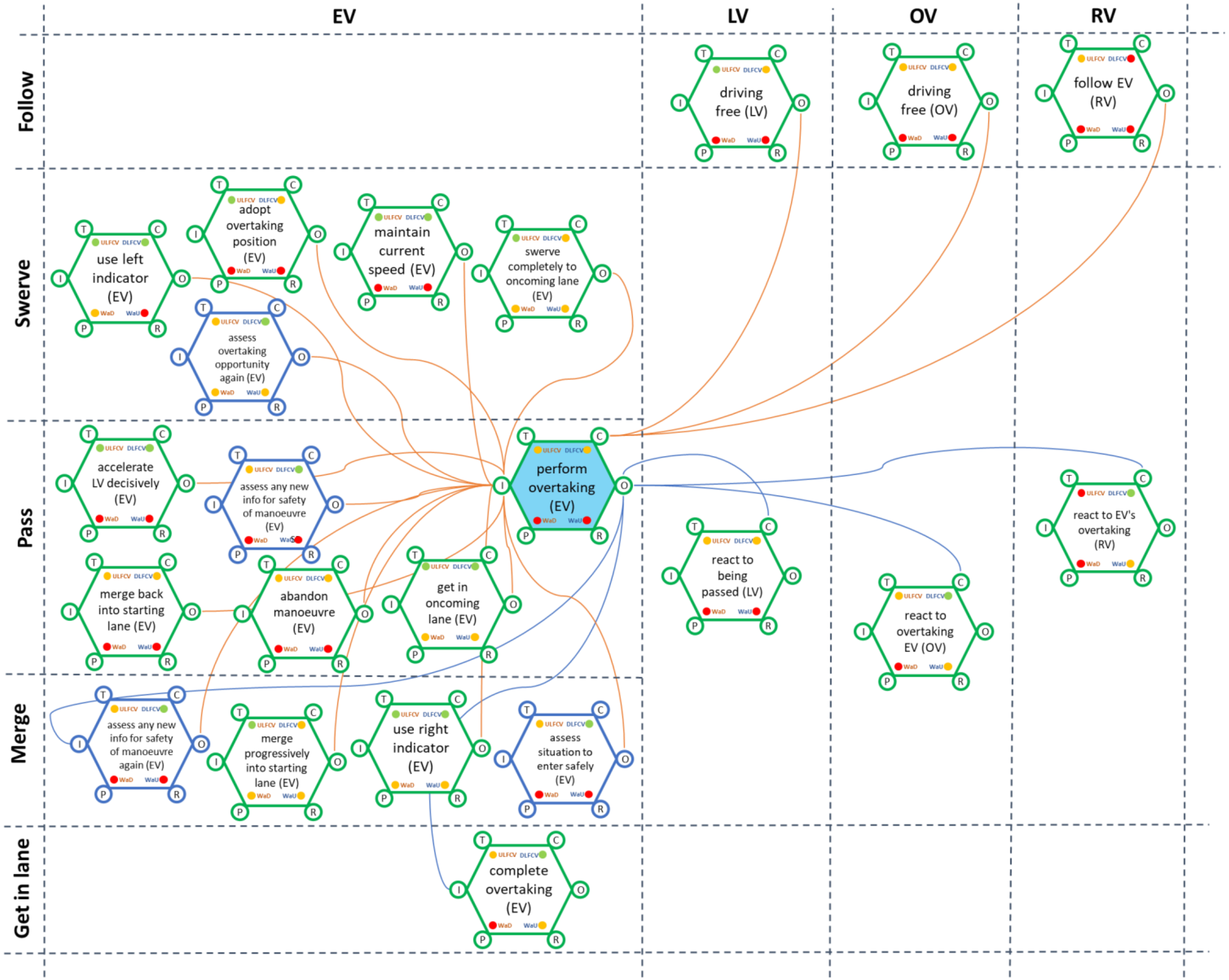

EV
LV
OV
RV
Follow
Swerve
Pass
Merge
Get in lane
driving free (LV)
driving free (OV)
follow EV (RV)
use left indicator (EV)
adopt overtaking position (EV)
maintain current speed (EV)
swerve completely to oncoming lane (EV)
assess overtaking opportunity again (EV)
accelerate LV decisively (EV)
assess any new info for safety of manoeuvre (EV)
perform overtaking (EV)
react to being passed (LV)
react to overtaking EV (OV)
react to EV's overtaking (RV)
merge back into starting lane (EV)
abandon manoeuvre (EV)
get in oncoming lane (EV)
assess any new info for safety of manoeuvre again (EV)
merge progressively into starting lane (EV)
use right indicator (EV)
assess situation to enter safely (EV)
complete overtaking (EV)
T
C
I
O
P
R
ULFCV
DLFCV
WaD
WaU

Figure B2. The critical path of the function “perform overtaking (EV)” in scenario five. It is the same for all other scenarios but with different values of the metrics.

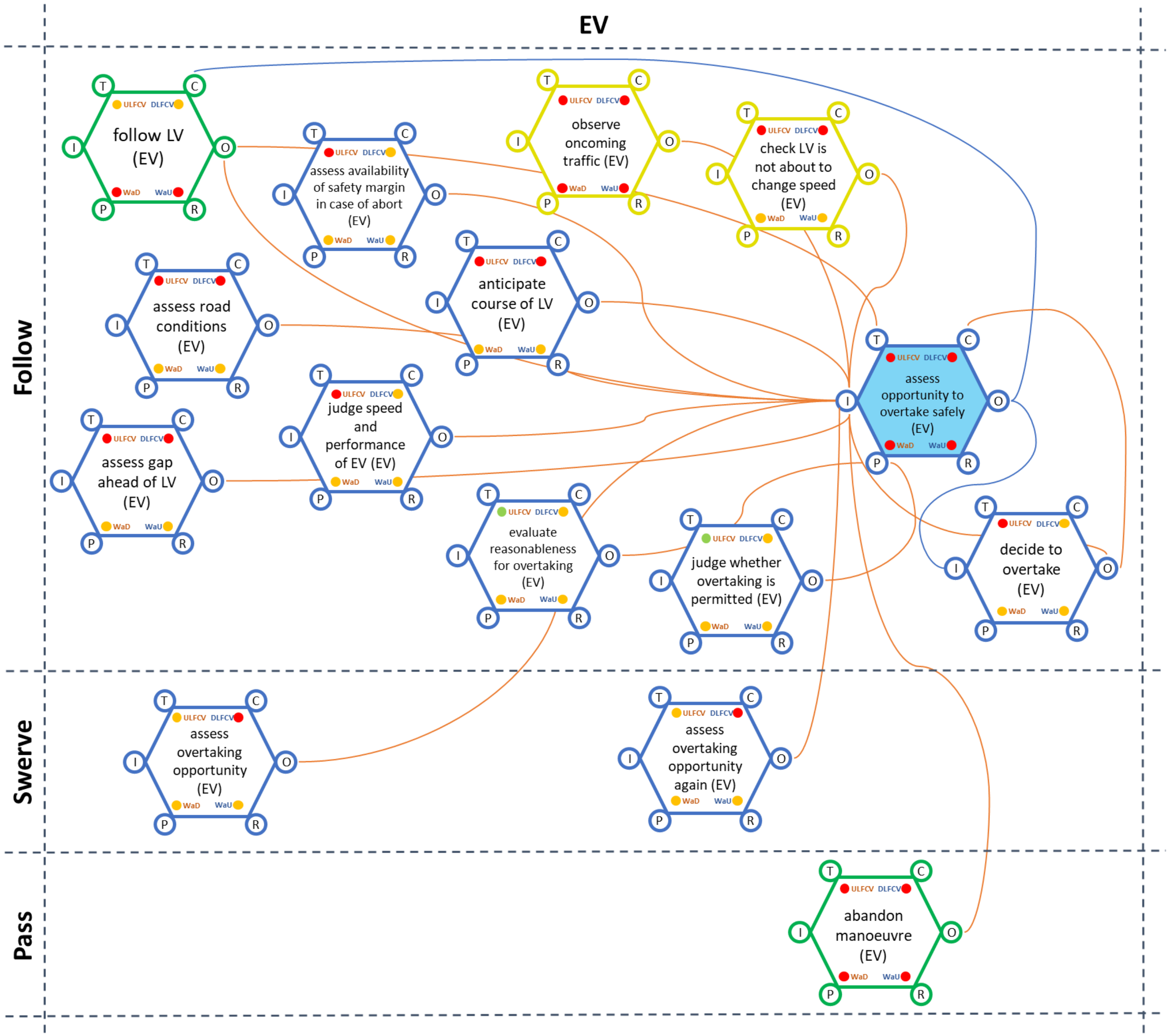

EV
Follow
Swerve
Pass
follow LV (EV)
assess availability of safety margin in case of abort (EV)
observe oncoming traffic (EV)
check LV is not about to change speed (EV)
assess road conditions (EV)
anticipate course of LV (EV)
judge speed and performance of EV (EV)
assess gap ahead of LV (EV)
evaluate reasonableness for overtaking (EV)
judge whether overtaking is permitted (EV)
assess opportunity to overtake safely (EV)
decide to overtake (EV)
assess overtaking opportunity (EV)
assess overtaking opportunity again (EV)
abandon manoeuvre (EV)
T
C
I
O
P
R
ULFCV
DLFCV
WaD
WaU

Figure B3. The critical path of the function "assess opportunity to overtake safely (EV)" in scenario two. It is the same for scenario four but with different values of the metrics.

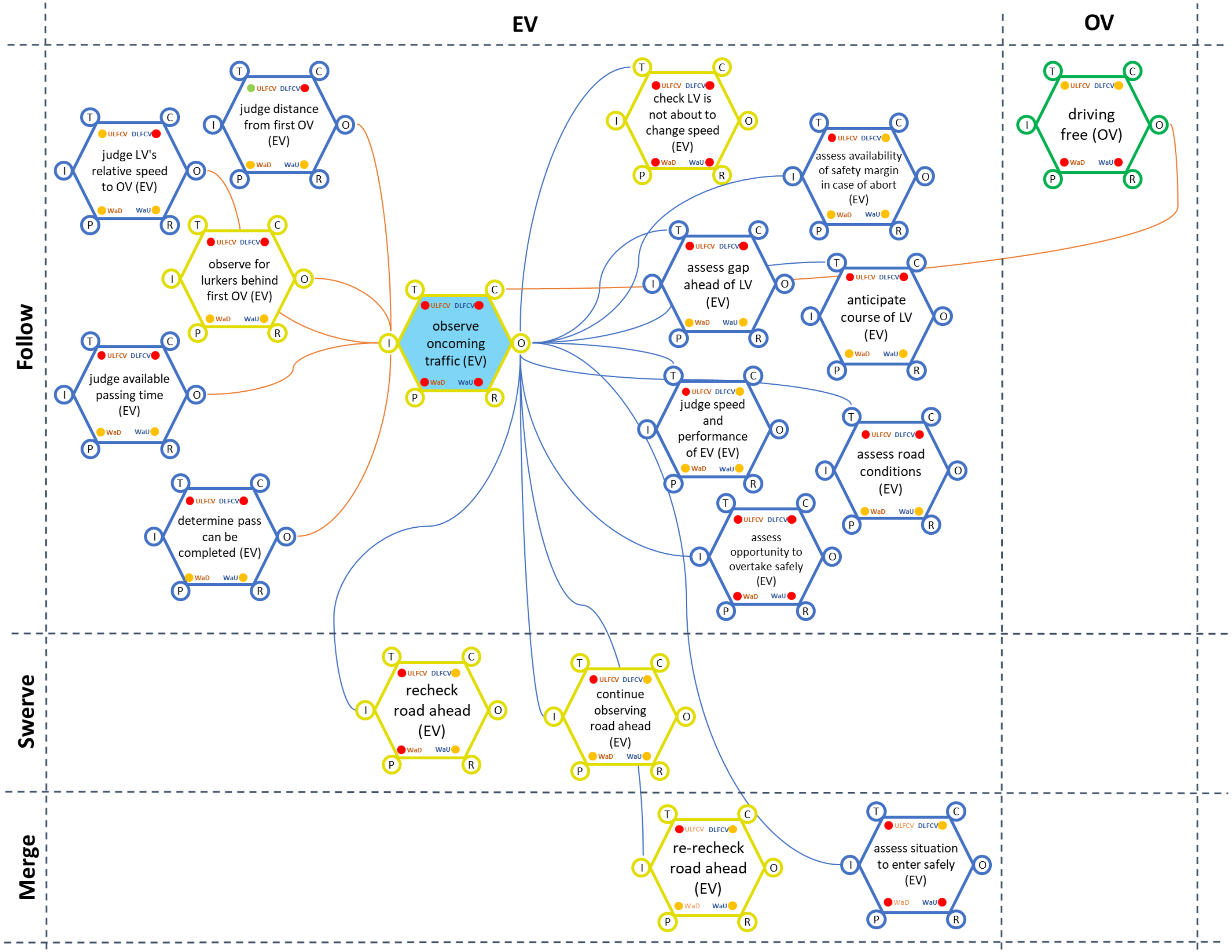

EV
OV
Follow
Swerve
Merge
judge distance from first OV (EV)
judge LV's relative speed to OV (EV)
observe for lurkers behind first OV (EV)
judge available passing time (EV)
determine pass can be completed (EV)
observe oncoming traffic (EV)
check LV is not about to change speed (EV)
assess availability of safety margin in case of abort (EV)
assess gap ahead of LV (EV)
anticipate course of LV (EV)
judge speed and performance of EV (EV)
assess road conditions (EV)
assess opportunity to overtake safely (EV)
driving free (OV)
recheck road ahead (EV)
continue observing road ahead (EV)
re-recheck road ahead (EV)
assess situation to enter safely (EV)
T
C
I
O
P
R
ULFCV
DLFCV
WaD
WaU

Figure B4. The critical path of the function “observe oncoming traffic (EV)” in scenario two. It is the same for scenario four but with different values of the metrics.

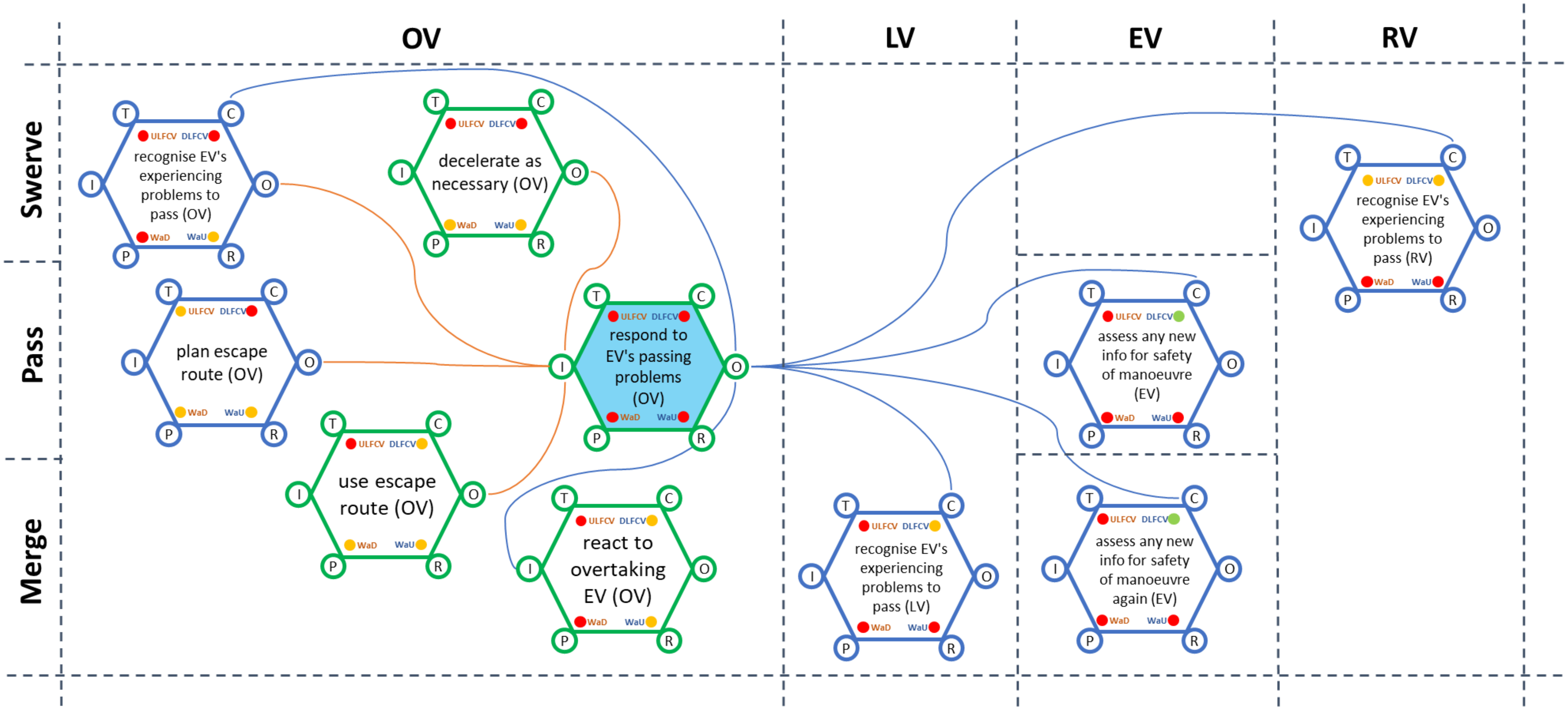

OV
LV
EV
RV
Swerve
Pass
Merge
recognise EV's experiencing problems to pass (OV)
decelerate as necessary (OV)
plan escape route (OV)
respond to EV's passing problems (OV)
use escape route (OV)
react to overtaking EV (OV)
recognise EV's experiencing problems to pass (LV)
assess any new info for safety of manoeuvre (EV)
assess any new info for safety of manoeuvre again (EV)
recognise EV's experiencing problems to pass (RV)
T
C
I
O
P
R
ULFCV
DLFCV
WaD
WaU

Figure B5. The critical path of the function "respond to EV's passing problems (OV)" in scenario three. It is the same for scenario four but with different values of the metrics.

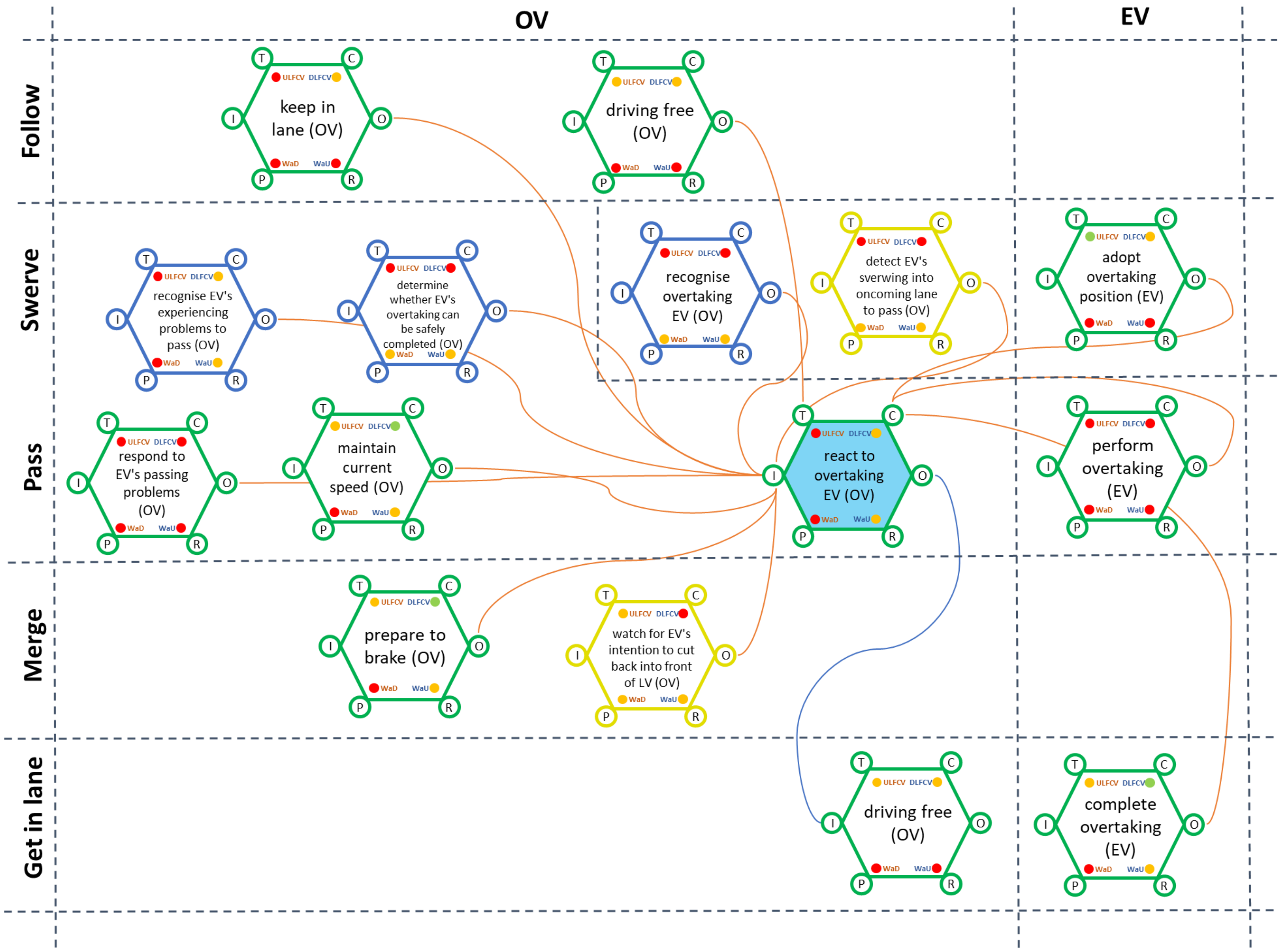

OV
EV
Follow
Swerve
Pass
Merge
Get in lane
keep in lane (OV)
driving free (OV)
recognise EV's experiencing problems to pass (OV)
determine whether EV's overtaking can be safely completed (OV)
recognise overtaking EV (OV)
detect EV's sverwing into oncoming lane to pass (OV)
adopt overtaking position (EV)
respond to EV's passing problems (OV)
maintain current speed (OV)
react to overtaking EV (OV)
perform overtaking (EV)
prepare to brake (OV)
watch for EV's intention to cut back into front of LV (OV)
driving free (OV)
complete overtaking (EV)
T
C
I
O
P
R
ULFCV
DLFCV
WaD
WaU

Figure B6. The critical path of the function "react to overtaking EV (OV)" in scenario four. It is the same for scenario three but with different values of the metrics.

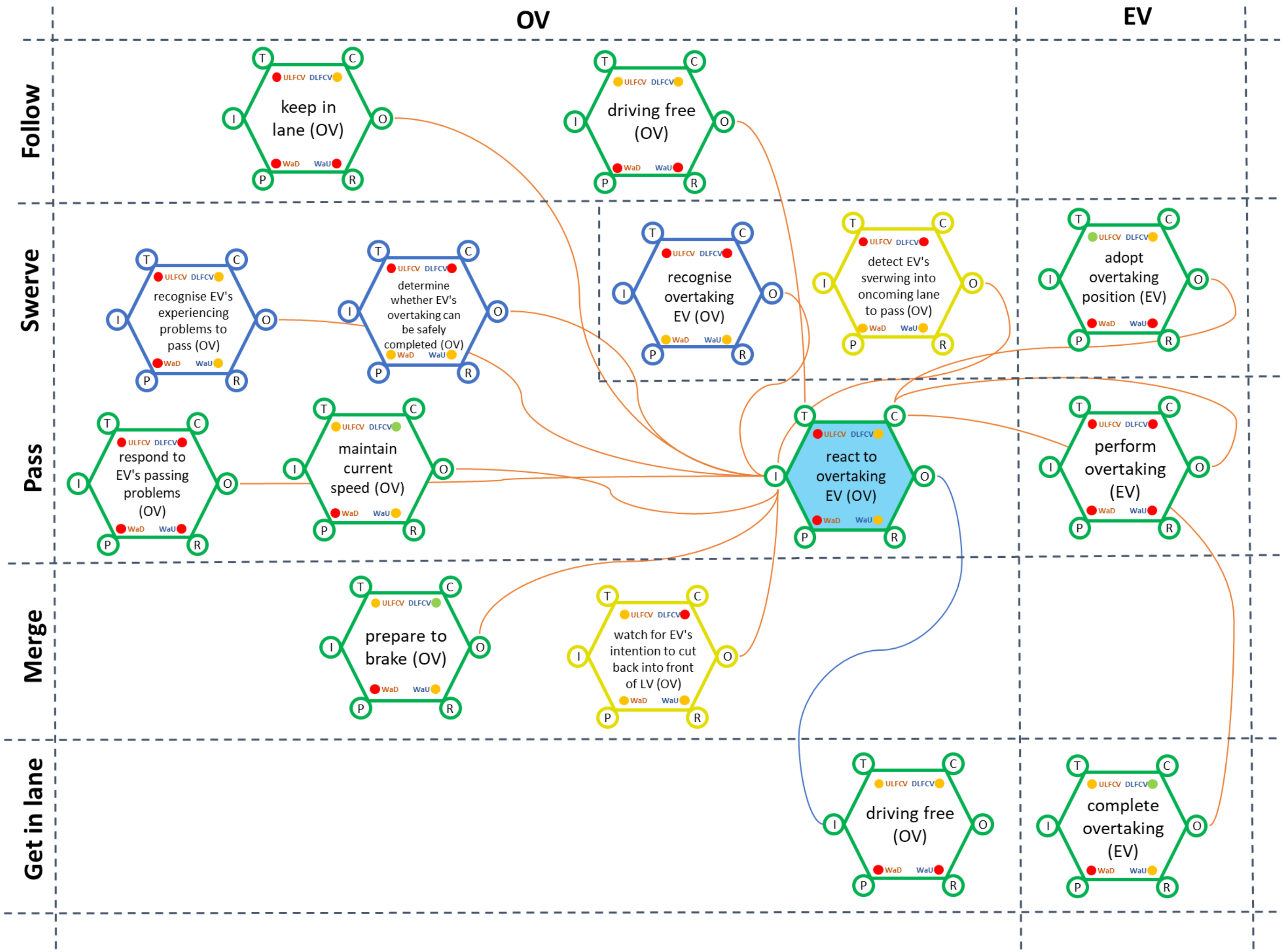

OV
EV
Follow
Swerve
Pass
Merge
Get in lane
keep in lane (OV)
driving free (OV)
recognise EV's experiencing problems to pass (OV)
determine whether EV's overtaking can be safely completed (OV)
recognise overtaking EV (OV)
detect EV's swerving into oncoming lane to pass (OV)
adopt overtaking position (EV)
respond to EV's passing problems (OV)
maintain current speed (OV)
react to overtaking EV (OV)
perform overtaking (EV)
prepare to brake (OV)
watch for EV's intention to cut back into front of LV (OV)
driving free (OV)
complete overtaking (EV)
T
C
I
O
P
R
ULFCV
DLFCV
WaD
WaU

Figure B7. The critical path of the function "recognise overtaking EV (OV)" in scenario three. It is the same for scenario four but with different values of the metrics.

B Article 1: “Safety of automated driving: the need for a systems approach and application of the Functional Resonance Analysis Method”

Grabbe, N., Kellnberger, A., Aydin, B., & Bengler, K. (2020). Safety of automated driving: the need for a systems approach and application of the Functional Resonance Analysis Method. *Safety science, 126*, 104665. https://doi.org/10.1016/j.ssci.2020.104665

Contents lists available at ScienceDirect

Safety Science

journal homepage: www.elsevier.com/locate/safety

Safety of automated driving: The need for a systems approach and application of the Functional Resonance Analysis Method

Niklas Grabbe*, Anna Kellnberger, Beyza Aydin, Klaus Bengler

Technical University of Munich, Chair of Ergonomics, Boltzmannstr. 15, 85748 Garching, Germany

ARTICLE INFO

Keywords:
Automated driving
Human driving
Risk assessment
Accident analysis methods and models
Systems thinking
Overtaking manoeuvre

ABSTRACT

Automated driving is technically advanced but proof of its safety is required for a successful market launch. Unfortunately, this evidence cannot be provided by current approval methods, something that is referred to as the so-called approval trap (Winner, 2015) and new test methods must be developed. This paper therefore argues in favour of the functional resonance analysis method (FRAM) as a risk assessment method in the development process of highly-automated vehicles, primarily to derive system design recommendations and secondly to provide essential insights into reducing the validation work. It begins with a systematic derivation of the benefits and suitability of FRAM. FRAM is then applied to an overtaking manoeuvre on a rural road in a road traffic case study to evaluate its suitability in more detail, followed by a discussion of the first application of FRAM to the road system and a presentation of its strengths as well as limitations. Finally, the conclusions consider the importance of the FRAM method in assessing risk and safety proactively for automated driving, also illustrating the need for further research.

1. Introduction

Traffic safety can be defined as the freedom from unacceptable risks and dangers in the change of location of persons or material assets (traffic objects) that are transported, for example, by the means of transportation from A to B. It includes the transport infrastructure and transport organisation (Schnieder & Schnieder, 2013, p. 74).

A basic distinction has to be made between two points of view in terms of safety (Schnieder & Schnieder, 2013, p. 67):

- protection of the environment from system impacts, which is referred to as safety
- protecting a system from external influences, which is called security

This paper only deals with the aspect of safety and not security.

According to Hughes et al. (2016), there are three main safety strategies or tools that can be applied to the traffic system components to enhance road safety. These are engineering, enforcement and education.

Engineering includes actions to avoid an accident or mitigate the damage of an accident. A distinction should be made between measures for the vehicle and for the infrastructure. Furthermore, countermeasures for the vehicle can differ with respect to primary safety and secondary safety. Primary safety comprises technical actions to avoid an accident. For instance, automated systems such as antilock braking system (ABS), electronic stability control (ESC) and adaptive cruise control (ACC) help the driver to cope with difficult driving situations. Secondary safety includes technical solutions to reduce the damage of unavoidable accidents. Examples are safety elements such as the crumple zone, safety belts and airbags. Design measures for the infrastructure can be the organisation of traffic through the layout of traffic roads, the presentation of information by road markings and traffic signs as well as protective functions such as crash barriers and the quality of the roads (Bubb et al., 2015, p. 57).

Enforcement is realised by traffic controls under the supervision by various institutions. The purpose is to make traffic participants aware of compliance with traffic rules and thus to bring this about and to punish their violation. Examples are speed measurements by radar, financial penalties or disqualification from driving.

Education describes interventions to improve the knowledge of necessary rules and their respective context, the ability to assess the traffic situation and skills for passive and active traffic participation. Examples are driving lessons, road safety education in school, driver safety training or targeted public campaigns (Schnieder & Schnieder, 2013, p. 446).

A further aspect could be the rescue chain in case of accident occurrence to reduce personal injuries and to save lives (Hughes et al.,

* Corresponding author.
E-mail address: n.grabbe@tum.de (N. Grabbe).

https://doi.org/10.1016/j.ssci.2020.104665
Received 19 August 2019; Received in revised form 5 February 2020; Accepted 6 February 2020

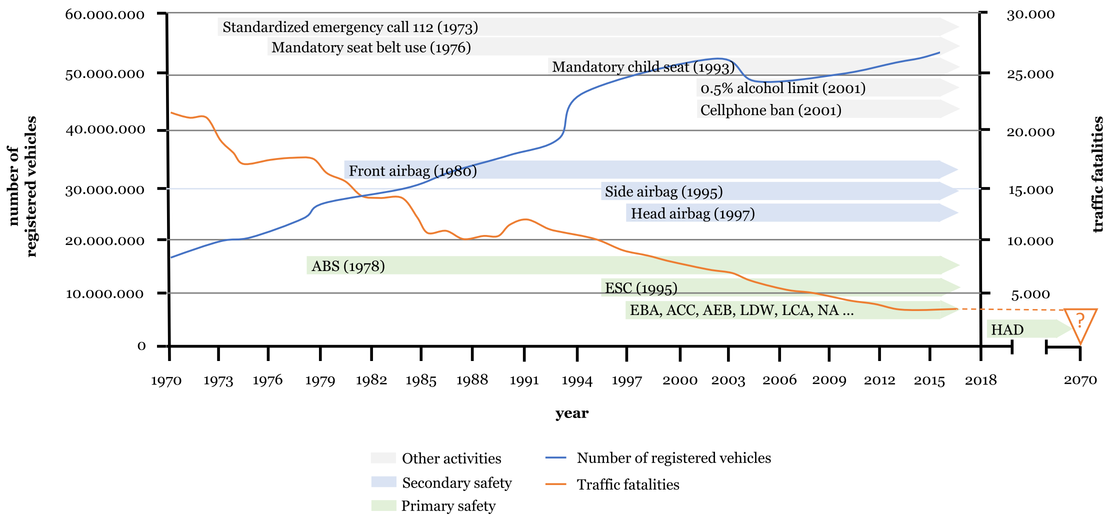


Fig. 1. Drop in traffic fatalities (orange line) due to safety improvements (horizontal boxes) despite the increase in registered motor vehicles (blue line) in Germany. The safety improvements belong to other activities (grey), secondary safety (blue) and primary safety (green), adapted from Winkle (2016a, p. 345). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

2016). The rescue chain means the idealized sequence in the treatment of accident victims in the context of first aid.

These safety strategies are combined in the past to improve the road safety and consequently the number of traffic fatalities dropped from over 21,000 in 1970 to almost 3000 a year today despite the huge increase in the number of motor vehicles registered in Germany (see Fig. 1). Fig. 1 illustrates the safety measures by the horizontal grey, blue and green boxes, for example the standardized emergency call 112 in 1973, the front airbag in 1980 or the ABS in 1978. It should be noted that the statistics for Germany only serves as an example for the subsequent explanations and arguments. Thus, it is also relevant for the settings of other countries.

Unfortunately, the number of fatalities has stagnated over the past years since 2012. The current approach and the stated countermeasure lie in the automation of the driving task with the long-term goal of fully automated driving. Thus, in the next years and decades, vehicles with SAE-level 3 or higher (HAD) according to SAE J3016 (2014) will be introduced. The underlying concept of the automation is in accordance with the safety strategy of engineering, specifically with the primary safety of the vehicle. It can also be attributed to infrastructure, if car-to-x communication (C2X) is taken into account. C2X is the radio-based exchange of information between vehicles, as well as between vehicles and other road users or in particular traffic infrastructure. This expands the coverage of vehicle sensors, such as radar or camera systems, by sharing the information of the sensors of other vehicles or the infrastructure.

The ACATech study (Lemmer, 2016) states that the introduction of automated vehicle guidance in addition to new mobility concepts can also lead to greater efficiency and safety. In general, automated driving should be safer, more efficient and comfortable (Maurer et al., 2016). The argument for increased automation in the driving task is often accompanied by the argument that the human in his role as a driver and main cause of accidents could be removed from the system. Consequently, the number of accidents would fall sharply. Similar quantifications of the safety potential of different advanced driver assistance systems (ADAS) can be found in Maier (2013).

However, this pattern of thinking follows a logic that is too simplistic. The concept of cause falsely links the logic of a clear causal link:

A causes B.

Thus, the driver would be the cause of the accident. If cause A is removed, then effect B disappears.

This relationship only applies to monocausal events. But a road accident is a rare, poisson-distributed and multi-causal event. Therefore, it is important to remember that the driver is involved in a road traffic accident as one of a number of factors and has not prevented this accident at the last moment:

A is involved in B in addition to other factors.

If A is removed, the other factors still apply for the remaining participants. Removing A would eliminate both the negative and positive effects that the human driver has on traffic. This results in a differentiated consideration of the mechanisms of accident development. Above all, is also addressed the mechanisms of accident prevention, in which drivers are as well involved in critical situations. In addition, the fact that not only is A (the driver) removed from the system but is replaced by X (the automated vehicle) with currently unknown effects, has to be considered.

All in all, the human driver is both an active or passive participant in an accident, as well as an accident avoidance and compensation element in the same system (Bengler et al., 2017).

Ultimately, technical solution must have an acceptable performance in situations that are critical for the human driver, but especially in situations that are not critical for the human driver. The work of Reichart (2000) gives an idea of a human's high performance in the subtasks of the driving task and shows that human errors occur with very low probabilities of 10^{-3} to 10^{-4} in the area of obscuring objects, interpretation or steering errors. These capabilities must be consistently achieved by the technical systems in the various traffic situations and constellations. Furthermore, Fastenmeier (2015) calculated that the human driver has a fatal accident every 90 million km. Considering that an average of 125 observations are made and 12 decisions taken every kilometre that is driven, these numbers show that a wrong decision leading to a fatal accident will be taken after about 10 billion observations and 1 billion decisions (cf. Huß, 1999). Shladover & Nowakowski (2017) made similar calculations for road traffic in the USA. One thing should have become clear from the calculation examples: it is a huge challenge for automation to achieve or even exceed the human driving performance.

We should also bear in mind the fact that not everyone benefits equally from automation. Both risk groups (Das et al., 2015; Visser et al., 2007) and accident black spots (Maier, 2013; Gründl, 2005), which represent a potential for automation, have been identified in literature. More recently, however, features relating to comfort in relatively safe scenarios (e.g. highways) are primarily addressed in the context of a market introduction and no orientation to specific groups of drivers or to human strengths and weaknesses is visible. Thus, in the worst case, we run the risk that the effect of automation will be zero and there will be no change in the accidents that occur (Bengler et al., 2017). Consequently, the technology could be without effect right at the beginning.

Finally, it can be concluded that valid proof of safety through automation is still pending due to the fact that current test methods are not economically and practically feasible for automated driving. This is the so-called approval-trap. Therefore, we have to create new test methods (Winner, 2015; Wachenfeld & Winner, 2016).

These new test methods should consider that the main goal must be to improve the safety of the entire system through the efficient interaction between humans, machines and other road users. Thus, the purpose of this paper is to identify and define an analysis method that can differentiate between the mechanisms of road traffic and identify the interdependencies between each element in the system. Besides, the method should be applied to specific and reasonable traffic scenarios so as to identify the contribution of the human driver to road traffic safety in these situations and to derive requirements for the automation and the potential of automation with its accompanying factors in these situations. Finally, if we place the method to be developed within the entire development process, we may be able to derive requirements and recommendations for the design of automated driving systems. Also, we can gain some useful information for the main emphasis of the validation process, such as criteria for exclusion, so as to reduce the validation effort (see Fig. 2). Here, Fig. 2 shows the development process for highly-automated vehicles as a V-Model. Over the course of development, the degree of abstraction of the system properties to be tested with the respective test method first decreases and then increases again. At the same time, the objective shifts from system design issues (left branch) to validation of intended properties (right branch). The method to be developed is located in the initial stage of the concept phase (orange oval), primarily to derive system design recommendations and secondarily to draw conclusions for the validation.

The paper is structured as follows. The second section argues why FRAM should be used to assess automated driving risk. This includes an overview of the development of accident analysis methods and models, safety perspectives as well as a description of the road traffic system. The third section defines the application of FRAM to the road system. It begins with an explanation of the structure of FRAM. This includes the basic principles as well as the various steps that must be taken in FRAM. The methodology of the application study as well as the results are then presented. The fourth section discusses the suitability of FRAM for a safety assessment of the road traffic system, especially regarding the effects of automated driving. Particularly, strengths and limitations are outlined. Finally, the conclusions anticipate the importance of the FRAM method to assess risk and safety proactively for automated driving, thus also illustrating the need for further research.

2. Finding a suitable method – Why FRAM?

In order to find a suitable method for the aim of this paper, the following subsections presents a brief overview of the historical development of accident and risk analysis methods as well as their underlying models and theories. Additionally, different safety perspectives are introduced and the properties of the road traffic system and associated requirements for analysis methods are described. Finally, an overview of the main systemic methods is provided and their suitability is discussed.

2.1. Development of accident causation theories and analysis techniques

A large number of accidents and incidents are analysed to gain new insights into the errors and their effects, often using a strict framework, such as a method, on which a model can be built or vice versa (Wienen et al., 2017).

There are several ways to classify these methods and models. The most commonly used classification system is into sequential, epidemiological and systemic analysis techniques (Qureshi, 2007; Wienen et al., 2017). As can be seen in Fig. 3, sequential techniques are by far

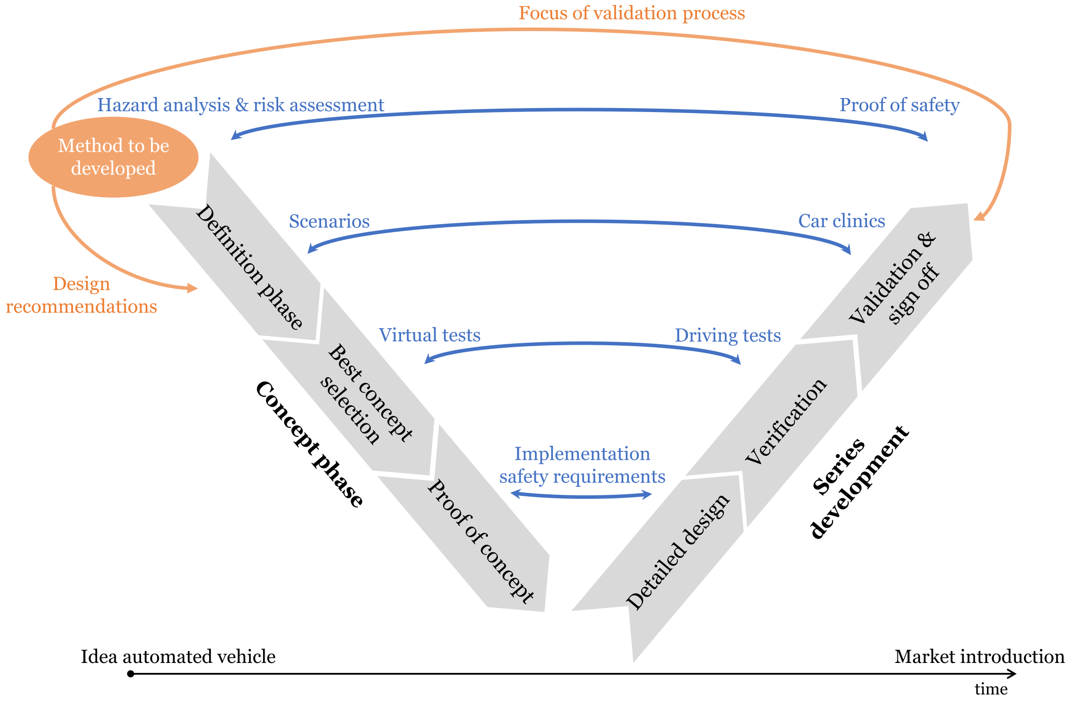


Fig. 2. Development process for highly-automated vehicles as a V-Model, adapted from Winkle (2016b, p. 608). The method to be developed is located in the top left corner (orange oval). Design recommendations should primarily be derived and secondary conclusions could be drawn for the validation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

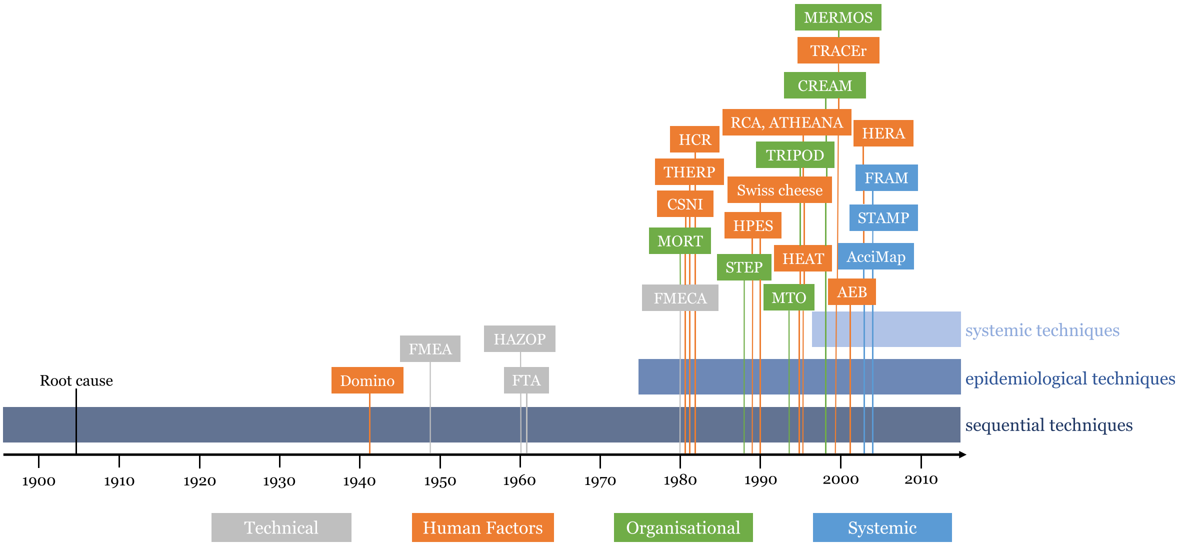


Fig. 3. Overview of the development of accident analysis techniques and important accident analysis and risk assessment methods, adapted from Underwood (2013, p. 18-19, 27) and Eurocontrol (2009). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the oldest, while systemic approaches only appeared 20 years ago. Another classification option for the accident analysis and risk assessment methods is to break these down into technical, human factors, organisational and systemic methods (Eurocontrol, 2009), see Fig. 3. Thus, the technical and systemic methods correspond to the sequential and systemic techniques of Underwood, whereas the organisational methods and approaches from the human factors are a mixture of sequential and epidemiological accident analyses.

The basic principles of the sequential, epidemiological and systemic analysis methods are shown below.

2.1.1. *Sequential accident analysis methods*

One of the first accident analysis theories, the Domino Theory, was introduced by Heinrich (1941). It describes an accident as a chain of discrete events occurring in a particular time sequence. The accident factors are arranged like dominoes one behind the other, in other words, each factor is dependent on the previous factors and accidents can be avoided by removing one of the preceding factors. This theory can be classified as one of the sequential accident analysis techniques, which include, among others, the failure or event tree analysis (FTA) and the failure mode and effects analysis (FMEA) (Leveson, 1995). These methods are only suitable for simple systems where losses are caused by failures or human error. They assume that the cause and effect relationship is linear and deterministic (Qureshi, 2007), where the causal network of events and states contains no feedback loops (Wienen et al., 2017). The advantage of these methods, however, is that they are already relatively old, and therefore mature, and they also provide an easily understandable, sequential picture of the events that lead to the accident (Wienen et al., 2017).

2.1.2. *Epidemiological accident analysis methods*

In the mid 1980′s, further factors or conditions and explanations were needed to understand the disaster at Chernobyl nuclear power plant or the loss of the space shuttle Challenger. The focus changed from human factor to organisation (Hollnagel, 2017).

New concepts and theories, such as Reasons (1990) Swiss cheese model, were thus developed. This model represents the different levels of safety as concatenated, rotating cheese slices, where the holes represent vulnerable points. In an unfavourable combination of these slices, for instance an interaction of several harmful factors, all holes fall in the same place and an accident occurs. However, if one layer is in the way, the hazard cannot develop into an accident. The cognitive reliability and error analysis method (CREAM) of Hollnagel (1998) for example, can also be listed here.

Overall, the epidemiological methods and their underlying models regard accidents as the result of a combination of different interacting factors, analogous to the spread of a disease (Qureshi, 2007). These factors are partly manifest and partly latent. Latent states are those that are present in the system long before the accident occurs, but are only recognized after the accident (Hollnagel, 1999).

The introduction of these factors improves an understanding of accidents, which contributes to the analysis of complex systems, but the causality is still linear and the links between states are loose, something that does not adequately represent the dynamics of a system (Hollnagel, 2004).

2.1.3. *Systemic accident analysis methods*

Since neither sequential nor epidemiological accident methods represent the dynamics and nonlinear interactions of complex sociotechnical systems and focus primarily on the “sharp end” factors (Dallat et al., 2017), new accident models had to be developed based on system theory (Leveson, 2004). The most widely-used systemic models are Leveson's (2004) systems-theoretical accident and process model (STAMP) and Hollnagel's (2004) first proposal of the functional resonance analysis model (FRAM). The strong connections between the various components of the system, which influence each other directly, are characteristic of systemic models and their derived methods. The methods try to describe performance at the level of the entire system and see the accident process as a complex, networked event that cannot be broken down into its individual components. Emerging events caused by complex interactions between the various system components can affect the performance of the system and cause an accident (Laaraj & Jawab, 2018; Qureshi, 2007; Wienen et al., 2017).

Since the systemic approach is of great relevance for the further course of this paper, the most important characteristics are summarized below (Laaraj & Jawab, 2018):

- emergence of a combination of interconnected and complex events
- macroscopic view of the system
- focus on the overall picture
- consideration of the effects of nonlinear interactions
- consideration of the complexity of the system
- consideration of the dynamics of the system.

2.2. Safety-I and Safety-II

Having considered the historical development of accident causation theories and analysis methods, we should also take a look at our comprehension of safety. The pertinent literature basically identifies two different ways of thinking about safety: the perspective of safety-I and safety-II.

Safety-I is described as a situation in which as few things as possible go wrong. The common assumptions are (Hollnagel, 2019):

- the system can be decomposed into meaningful elements
- the function of each element is bimodal (true/false, work/fail)
- the failure probability of elements can be analysed individually
- the order of events is predetermined and fixed

This point of view evolved in the 1920s, where systems were loosely coupled, linear and stable and system functions were easy to understand and completely describable (Hollnagel, 2018). Finally, the whole system is seen as equal to the sum of its parts. However, most of current systems are tightly coupled, increasingly non-linear, less stable and system functions are hard to understand due to their complexity. In such complex systems the whole is greater than the sum of its parts and outcomes cannot be totally controlled or predicted. Thus, the perspective changed to safety-II.

Safety-II can be seen as a situation in which as many things as possible go right. The basic assumptions are (Hollnagel, 2019):

- systems cannot be understood by decomposing them
- functions are not bimodal, in fact performance is always variable
- this performance variability is a source of success as well as of failure
- the functions must be flexible to fit the conditions

Finally, from the point of view of safety-I, the human is seen as a hazard and performance variability should be prevented, whereas the safety-II perspective regards humans as an inevitable resource for system flexibility and resilience and performance variability should be monitored and managed. Here, a critical perspective on the two safety views in relation to automated driving and road safety unveils one interesting fact: the point of view of safety-I supports the argumentation of the replacement of the human driver by automation and on the opposite the point of view of safety-II argues in favour that the human driver is still necessary in at least some situations due to system flexibility and that current automated systems probably are not able to cope with this flexibility (cf. De Winter and Hancock, 2015) in any situation. This indicates that the motivation of full automation and the associated goal of increased traffic safety is largely safety-I driven and the safety-II perspective is almost neglected. In fact, this is also reflected in the aforementioned definition of traffic safety in the introduction, which is largely safety-I oriented. Thus, the application of the safety-II perspective on automated driving seems urgently required. In particular, this approach could be the answer to the outstanding issue of what can be opposed to the previous thinking in terms of the bimodality "right or wrong" (Winner, 2015) to overcome the approval-trap and the statistical approach which is based on track distance that currently focuses only on accidents.

However, adopting the safety-II perspective does not mean that the safety-I approaches and techniques used up to now have to be completely replaced: rather we should look at what is happening differently. Although the two perspectives differ in many respects, they represent two complementary views of safety and underlying methods can also include both perspectives, not only one (Hollnagel, 2018).

Which methods and models are suitable ultimately depends on the system and events being analysed. Thus, one should first become aware of the system under examination and the aim of the analysis.

2.3. Road traffic system as a complex socio-technical system

The way we think about systems and their behaviour influences the methods and models we choose to manage safety and in more general to solve problems they pose. Thus, the system to be analysed, in other words the road system, has to be described, before we can trust a suitable type of automated driving risk assessment method or model. A closer look should be taken at the definition of a socio-technical system as well as the properties of complex systems for a better understanding of the road traffic system.

Socio-technical systems can be defined as the increasingly common classification of large systems that have a combination of technological systems (hardware and software), human interfaces, and organisational systems (Jackson, 2009). In such systems, requirements arise from interaction with the external environment as well as social, organisational, and individual factors within the system. These requirements must be met with limited resources. Variability in system performance is a feature of large socio-technical systems and makes a complete description of the work system confusing or impossible (Frost & Mo, 2014; Hollnagel, 2017).

According to Vicente (1999), a system comprising technical, psychological and social elements can be termed socio-technical. In this sense, Salmon et al. (2012) conclude that the road system, which connects all three elements for the purpose of transporting people and goods from one place to another, is of a socio-technical nature.

According to Dekker et al. (2011), the properties of complexity can be summarized as follows: complex systems are only held together by local relationships. No component is aware of the behaviour of the system as a whole and no one knows the full impact of their actions. The components react locally to the information available to them. The complexity comes from the huge networks of relationships and interactions that result from these local actions. The boundaries of what makes up the system become blurred, and the interdependencies and interactions multiply and spread quickly.

Salmon et al. (2012) demonstrated that the road traffic system is just such a complex system because it fulfils all of the aforementioned prerequisite properties. Specifically, these are the following factors:

- road systems are open systems due to influences from the environment, but also influences on the environment in return
- components are ignorant of the behaviour of the system as a whole
- no component achieves the level of complexity of the entire road system
- inputs need to be made by components at all times in order to keep the system functioning
- path dependence: previous decisions and actions influence the present time
- non-linear interactions: asymmetry between input and output

Consequently, the road transport system is a complex socio-technical system and could be embedded in the systemic quadrant of the system interaction-coupling matrix adapted from Perrow (1984), see Fig. 4. This means that sequential and epidemiological analysis techniques are inadequate and that an approach based on systemic methods should be used to best represent a risk assessment of automated driving. This is also confirmed by Larsson et al. (2010), who describe system theory as an important basis for safety work in complex socio-technical systems, such as the road traffic system. In addition, Hughes et al. (2015) examined 121 different models relevant for road safety in their work, divided them into seven different types and then evaluated them. They conclude that systemic models are best suited in both research and practice for application to the traffic system. Furthermore, Reichart (2000) investigated human reliability in driving motor vehicles and concluded that only a systemic view of the driver, vehicle and driving environment can capture the interaction of the traffic elements and effects in the traffic system adequately.

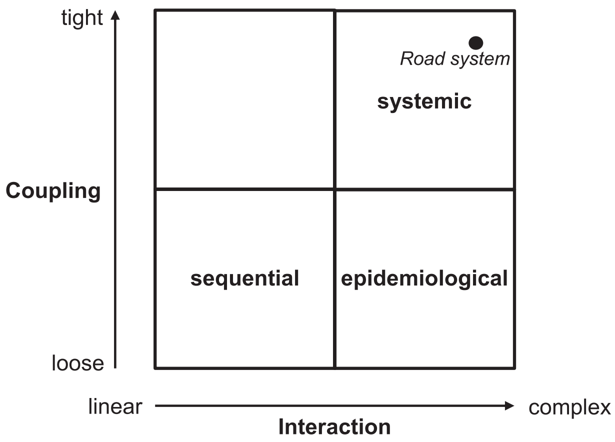


Fig. 4. System interaction-coupling matrix combined with accident analysis methods classification and assignment of the road system, adapted from Perrow (1984) and Wienen et al. (2017, p. 22).

Although researchers believe that the new system approach can overcome many of the limitations of traditional methods, only a few results are found in a literature review of system theory and traffic safety (Larsson et al., 2010), which emphasise the importance of this paper.

2.4. Comparison of systemic methods

As the systemic accident analysis and risk assessment methods seem to be the most appropriate to map the effects of road traffic automation in comparison with the human driver, an overview of the most commonly used systemic accident analysis methods, AcciMap, STAMP and FRAM, (Underwood, 2013) will be provided. The most important basics of these methods are explained and existing applications for road traffic are briefly described. Finally, we discuss their suitability for application to the road traffic system regarding a safety evaluation of automation.

2.4.1. AcciMap

AcciMap is an extension of the Risk Management Framework (Rasmussen, 1997) of Svedung & Rasmussen (2002) that is based on the idea that safety can be influenced by various decisions at all levels of the system hierarchy. The tool has therefore been designed to perform a vertical analysis of a socio-technical system for a given accident, and thus includes a study of events, actions, agents, and decisions that contributed to the accident (Underwood, 2013).

Traditionally, AcciMap is intended as use after accident occurrence. The authors do not think that this necessarily excludes the use of AcciMap for proactive risk assessment since a risk assessment can include data of both accidents and accident-free driving.

AcciMap has already been used to model the transport system: Young and Salmon (2015) used the Risk Management Framework to analyse the distractions of driving with regard to the responsible actors at different levels of the road system. Scott-Parker et al. (2015) used an AcciMap to demonstrate knowledge about the safety of young drivers in terms of the various actors, contributing factors and countermeasures. McIlroy et al. (2018) present a holistic view of the British road traffic system, illustrating the complexity of such a system in an ActorMap. Finally, Stanton et al. (2019) presented an AcciMap of the Uber collision with a pedestrian and recommend it as an appropriate approach to support road traffic collision investigations.

However, the limited availability of accident analysis data, especially as regards the influences of higher hierarchy levels and the degree of subjectivity in the categorisation, is generally criticised.

2.4.2. STAMP

STAMP (Leveson, 2004) is an accident model in which safety is treated as a dynamic control problem, not as a reliability and failure problem. Safety is controlled by a control structure embedded in an adaptive socio-technical system. The control structure model views systems as interconnected components that maintain a state of dynamic equilibrium through feedback loops of control and information. An accident is the result of a loss of control or inadequate enforcement of safety-related restrictions in the development, design or operation of the system (Leveson, 2004; Leveson, 2011).

In addition, there are two methods that are based on STAMP: on the one hand the Systems-Theoretic Process Analysis (STPA) method, which is used in hazard analysis to analyse possible causes of future accidents, and on the other hand the Causal Analysis based on STAMP (CAST) which is used to describe and understand an accident (Leveson, 2011).

STPA uses a feedback loop safety control structure diagram to identify unsafe scenarios and develop detailed safety constraints. STPA can be performed within four main steps: to define the purpose of analysis, to build a model of the system as a control structure, to define unsafe control actions and to identify possible reasons for unsafe control which is done by creating scenarios (Leveson & Thomas, 2018).

Some examples for the application of STPA in the context of automated driving can be found in Abdulkhaleq et al. (2018, 2017). They conclude that STPA is a useful approach for identifying more types of detailed requirements in addition to the safety-in-use (SIU) requirements. SIU is defined complementary to functional safety as absence of unreasonable risk due to hazards not caused by malfunctioning. Also, STAMP has already been applied in road safety to identify the gap in the control structure of the road network (Salmon et al., 2016). The goal of the analysis was to explore the range of actors and organisations within the road system in Queensland and their key relationships. Here, the STAMP methodology has certain limitations. A number of the identified mechanisms are not controls but can be described as influencing mechanisms. Such an extension of the method would allow not only the development of a control structure but also the creation of an influence structure. Moreover, STAMP is unable to adequately represent the broader societal impact on road users and the behaviour of the transport system. This argument is consistent with the observation that STAMP often cannot adequately account for the environmental conditions of the system. Furthermore, one criticism is that control and feedback loops can only be described between different levels and not within one level. In addition, Alvarez (2017) analysed the safety benefits of automated driving using a STAMP-based approach. The author concluded that the conceptual framework is suitable, but that other system theory approaches such as the Risk Management Framework or FRAM may also provide an appropriate conceptual framework. Unfortunately, a comparison between the three approaches was beyond the objectives of their work, so the question remains unanswered as to what the differences are between these methods in terms of safety assessment of automated driving.

2.4.3. FRAM

FRAM (Hollnagel, 2017) began as a qualitative method for risk assessment and accident analysis. It allows the modelling of complex socio-technical factors, including their interfaces between human and technology, coupling and dependency effects, nonlinear dependencies between subsystems, and functional variability (Woltjer & Hollnagel, 2008). FRAM is a method to analyse how something happens or how a system works and to generate a model of that. The objective is to describe and understand socio-technical systems in terms of functions rather than in terms of components. A FRAM model represents the functions as work-as-done and focus on adjustments of everyday performance which usually contributes to things going right. If these performance adjustments aggregate in unexpected ways, functional resonance will occur and accidents are the result. The final purpose is not to eliminate performance variability but to investigate and to monitor what is necessary for everyday performance to go right, trying to dampen variability in order to reduce resonance effects and unwanted outcomes (Hollnagel, 2017). The basic principle is that you need to understand how work is done when it goes well in order to understand what happened when it failed.

FRAM has been used in many different fields, including aerospace (Hollnagel et al, 2008; De Carvalho, 2011), nuclear power (Lundblad et al., 2008), the oil industry (Shirali, et al., 2013) and rail transport (Steen and Aven, 2011; Belmonte et al., 2011). Ferreira and Cañas (2019) in particular investigated the potential impacts of automation on air traffic control operations by exploring how the interactions between human operators and technology may change if new automation features are introduced into the system. This represents a similar application compared to the background of this paper (the automation of the driving task).

However, to the best of our knowledge, a FRAM analysis of the road system has not yet been published by other researchers. Smoczyński et al. (2018) noted that the published papers and the description of FRAM in other fields of application, particularly traffic-related applications such as maritime (Patriarca & Bergström, 2017) or air transport (Yang, Tian & Zhao, 2017), indicate the possibility of applying the method to the road traffic system.

In general, FRAM is seen as a useful tool to build an understanding of a system actual mechanisms and workings that are needed to initiate a learning process and to support the risk management activities concerning the proactive assessment of technological changes and their impacts (Ferreira and Cañas, 2019). Typical results of a FRAM analysis are that it contributes to the understanding of real work and unveils unsafe functional interactions within one agent and between different agents that are often underestimated by traditional methods and design approaches (Patriarca & Bergström, 2017; Ferreira and Cañas, 2019).

This applies to the road system which has in particular many non-linear dependencies between vehicles, drivers, vulnerable road users, infrastructure elements and environmental conditions. Despite this high level of complexity and the occasionally simultaneous high-risk behaviour of some road users, accidents are rarely the result due to high adaptability or resilience of the system. Therefore, it is expected that FRAM can identify these resilience mechanisms of the road system and even assess the impact of introduced automation.

2.4.4. Discussion of the suitability of systemic methods

Underwood (2013) compares the three methods in terms of their fulfilment of criteria of the system approach. Fig. 5 shows this comparison and the first thing that becomes clear is that FRAM is the only method that fulfils all criteria, STAMP cannot fulfil equi- and multi-finality, and AcciMap cannot fulfil goal seeking and equi- and multi-finality. However, equi- and multifinality play an important role in the road system because the interaction of different human components with their behavioural variations leads to many different outcomes with the same inputs or many different developmental paths may have a similar result. Furthermore, it can be seen that FRAM is the method that explicitly identifies most features. Thus, FRAM offers the greatest potential for safety analysis in road traffic according to Underwood's assessment criteria.

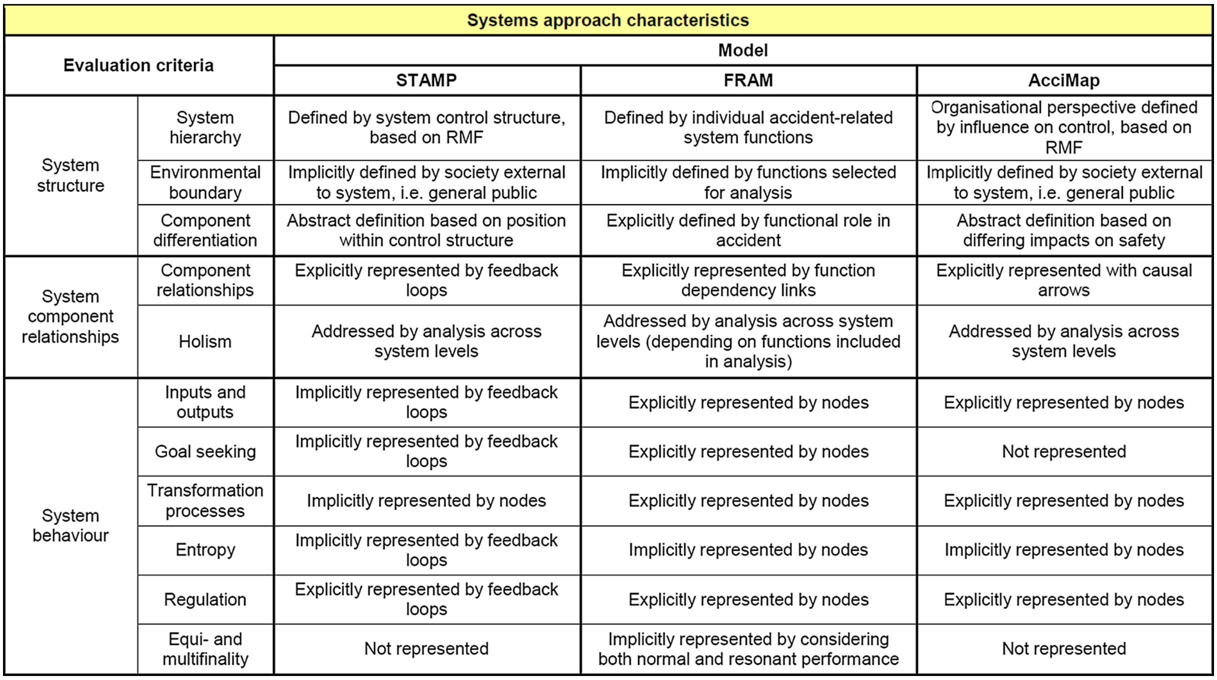
Additionally, considering the recent developments of these methods over the last decade, STAMP and Accimap have remained methodologically the same, while the traditional approach of FRAM has recently been significantly supplemented by numerous extensions. The most important research and extensions are listed below: the use of FRAM in combination with Rasmussen's abstraction hierarchy by Patriarca et al. (2017a) enables on the one hand in addition to the functional system description a representation of the system structure and its hierarchies and on the other hand a better handling of the complexity. Another opportunity to reduce the complexity is shown through the transformation of a FRAM model into a matrix representation by Patriarca et al. (2018). In particular, the quantification approach of Patriarca et al. (2017b) significantly enhances the basically qualitative and subjective nature of analysis. A further example to support and enhance the traditional qualitative inputs by objective inputs from data of sensor technologies is represented by Arenius (2017). Moreover, the introduction of the new software tool myFRAM (Patriarca et al., 2017c) contributes to a more standardised and systematic implementation of a FRAM analysis and also offers an interface for combining FRAM with many other methods. Last but not least, the suggestion of Belmonte et al. (2011), that predictive models can be calibrated by inputting data from real-world or simulated scenarios to test the internal validity of a FRAM model, could be facilitated by the approaches of Tian et al. (2016) and Slater (2016). The aforementioned examples emphasise the great potential of FRAM.

It should also be pointed out that AcciMap and STAMP are model-cum-method approaches. This means that these approaches have an underlying model that defines a set of relations and that the associated methods offer an interpretation of events in terms of these relations. This imposes an a priori knowledge of the structure of an event. Thus, the assumptions of a model must be correct, otherwise the results of an analysis will be useless. This does not match the perspective of safety-II (but this is required, cf. Section 2.2), which implies that complex socio-technical systems are not fully understandable. Instead, FRAM can best be described as doing the opposite. FRAM describes systems in terms of functions without any predefinition of specific functions or assumptions of organisational structure. In addition, relations between functions are defined by empirically determined functional dependencies rather than by hypotheses of the underlying model. Moreover, FRAM does not belong to any model or any assumptions about possible causes and cause-effect relationships. Hence, FRAM is a method-sine-model approach due to the use of a method to produce a model and not vice versa (Hollnagel, 2017).

Furthermore, FRAM is scale invariant, whereas model-cum-methods are scale variant. The advantage of scale invariance is the simplicity of the method. It needs no large taxonomies or explanations that can be cumbersome to use or constrain the depth and breadth of an analysis (Hollnagel, 2017).

Finally, we should also bear in mind the fact that FRAM models the system at functional level, whereas STAMP provides a view at component structure level. This means that STAMP needs a complete system architecture for risk identification, which makes assumptions about the system and its mechanisms. Instead, FRAM does not make any assumptions about the processes and the involved components. Thus, the principle of STAMP contradicts with the stated goal of this work, where the mechanisms of the system must first be revealed and understood. Here FRAM is significantly more appropriate because it contributes to what a system does and not what it is. We therefore recommend the use of FRAM at the very beginning of a product development cycle (see Fig. 2) and STAMP at a subsequent position, for instance when a solution has been identified and implemented for the system, to especially derive risks at the component level. Another opportunity is the meaningful combination of both methods, for example the integration of components in a FRAM model, which could be done by the use of the abstraction/agency framework presented by Patriarca et al. (2017a).

In the end, it can be concluded that FRAM is the most promising method to achieve the goal of the paper as explained in the introduction in Section 1: to find an analysis method that can differentiate between the mechanisms of road traffic and identify the interdependencies between each system element. This method should be applied to specific and reasonable traffic scenarios in particular to identify the contribution of the human driver to road traffic safety within these situations and to derive requirements for the automation and the potential of automation in these situations with its accompanying factors. Thus, the next section deals with the applicability of FRAM in a case illustration. Nevertheless, it would be reasonable also to investigate other systemic methods, especially STAMP/STPA, in a similar case study in future research in order to allow the best possible comparison of the methods in terms of safety assessment of automated driving. However, this is



Systems approach characteristics				
Evaluation criteria		Model		
		STAMP	FRAM	AcciMap
System structure	System hierarchy	Defined by system control structure, based on RMF	Defined by individual accident-related system functions	Organisational perspective defined by influence on control, based on RMF
	Environmental boundary	Implicitly defined by society external to system, i.e. general public	Implicitly defined by functions selected for analysis	Implicitly defined by society external to system, i.e. general public
	Component differentiation	Abstract definition based on position within control structure	Explicitly defined by functional role in accident	Abstract definition based on differing impacts on safety
System component relationships	Component relationships	Explicitly represented by feedback loops	Explicitly represented by function dependency links	Explicitly represented with causal arrows
	Holism	Addressed by analysis across system levels	Addressed by analysis across system levels (depending on functions included in analysis)	Addressed by analysis across system levels
System behaviour	Inputs and outputs	Implicitly represented by feedback loops	Explicitly represented by nodes	Explicitly represented by nodes
	Goal seeking	Implicitly represented by feedback loops	Explicitly represented by nodes	Not represented
	Transformation processes	Implicitly represented by nodes	Explicitly represented by nodes	Explicitly represented by nodes
	Entropy	Implicitly represented by feedback loops	Implicitly represented by nodes	Implicitly represented by nodes
	Regulation	Explicitly represented by feedback loops	Explicitly represented by nodes	Explicitly represented by nodes
	Equi- and multifinality	Not represented	Implicitly represented by considering both normal and resonant performance	Not represented

Fig. 5. Evaluation of systems approach characteristics of STAMP, FRAM and AcciMap (Underwood, 2013, p. 71).

beyond the scope of this paper.

3. Application of the Functional Resonance Analysis Method

In a next step, FRAM was applied to road traffic in a case study to assess the suitability of FRAM in more detail. We will begin by explaining the structure of FRAM. This includes the basic principles as well as the various steps that must be taken in FRAM. Finally, the methodology of the application study as well as the results are presented.

3.1. The FRAM structure

3.1.1. Basic principles

The original FRAM relies on four principles, which can be seen as a summary of the experiences gained in safety management in the past and is also associated with the disability of traditional safety methods to cope with the properties of complex socio-technical systems (Hollnagel, 2017):

- Equivalence of failures and successes:
  Failure and success spring from the same source, i.e. everyday work variability. This leads both to things going right, as they should, but sometimes also causes things to go wrong. Thus, success and failure are not of a different nature, indeed things go right and wrong for the same reasons.
- Approximate adjustments:
  Socio-technical systems are partly intractable and work conditions are underspecified. Thus, resources and time are usually limited and sometimes insufficient. Therefore, humans, individually or collectively, and organisations adjust their everyday performance to match the situation. This is also called the Efficiency-Thoroughness Trade-Off (ETTO) principle (Hollnagel, 2009). Moreover, this is an additional argument for the principle of equivalence described above.
- Emergence:
  However, it is not possible to explain things that are going happen as resultant for an increasing number of systems or events. In fact, the outcome is said to be emergent. This means that causes cannot be explained by principles of decomposition and causality, but rather outcomes may be due to a particular combination of transient conditions that only were present at a particular point in time and space without leaving any traces. This means that effects are non-linear and causes have to be reconstructed rather than found. A more powerful explanation for emergent outcomes is ultimately required that leads to the fourth principle.
- Functional resonance:
  Functional resonance is the detectable signal that emerges from the unintended combination of the variability of many signals. This resonance is not stochastic or random, but rather more systematic due to certain regularities which are characteristic for different functions. Therefore, functional resonance explains both emergent and non-linear outcomes to enable their predictability and control. Thus, safety analyses can be based on the presence of variability.

3.1.2. Step 0: Recognise the purpose of analysis

In a first step 0, before the actual method begins, practitioners of FRAM have to make the purpose of using FRAM clear. A FRAM model can be used to understand how an event happened (retrospective event analysis), to assess how something may happen (prospective risk assessment) or evaluate the effects of measures to improve system design (new or redesigned systems).

3.1.3. Step 1: Identification and description of a system's functions

The first step in FRAM is to identify the functions that are essential for the success of everyday work. A function refers to the tasks (work-as-imagined, WAI) or activities (work-as-done, WAD) that have to be done to produce a certain outcome. Each function is characterised by six different aspects (Hollnagel, 2017):

- Input (I): energy, matter or information which is used or transformed by the function to produce an output or what activates or starts a function.
- Output (O): the result of what a function does, either an entity or state change.
- Precondition (P): conditions that must be fulfilled to carry out a function, but a precondition itself does not work as a signal that starts a function.

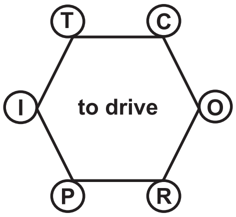


Fig. 6. Example of a graphical representation of a FRAM function.

- Resource (R): is what a function needs or consumes while the function is carried out (e.g. matter, energy, information, software, tools and operators).
- Control (C): is what monitors or regulates a function to produce a desired outcome (e.g. plans, procedures, guidelines or a software program)
- Time (T): temporal constraints of the function in terms of both duration and time of execution

A FRAM function with the six aspects is traditionally represented graphically by a hexagon (see Fig. 6). The software Functional Model Visualiser (FMV) (Hill & Hollnagel, 2016), available at http://functionalresonance.com/support/software%20development.html, or its pro-version designated as FMV Pro, available at http://www.zerprize.co.nz/FRAM/index.html, should be used to describe the functions and to generate a graphical model.

Furthermore, it is helpful to divide the functions into two classes: foreground and background functions. Foreground functions are the core of the analysis and may vary significantly during an instantiation of the model. Background functions represent common conditions that are relevant for and used by foreground functions. They are outside the scope of analysis and do not vary significantly. Background functions can be seen as the boundary of the model.

Event reports, procedures, design specifications, story-telling, field observations or interviews can be used to start making a list of essential functions. Any function can be used to start the FRAM model before proceeding in an iterative manner to identify the remaining functions and their couplings. These couplings represent a specific instantiation of a model. In this process, functions have temporal and causal relationships, thus functions that happen before other functions and affect them are called upstream functions whereas, functions that happen after other functions and are affected by them are called downstream functions.

The model can be refined through discussions and iterations, and it is completed if no aspect occurs for one function only, so that all aspects defined for one function have to be included in the aspects of other functions. The stop rule for system boundary is semi-explicit, namely that the analysis should stop if no unexplained variability of functions remains (Hollnagel, 2017, p. 59-60). Finally, the model can be calibrated using so-called subject matter experts (SME's).

3.1.4. Step 2: Identification of performance variability

In a second step, the objective is to identify and characterise the variability of functions. The identification of variability is crucial to understand potential as well as actual couplings between functions, which can lead to unwanted outcomes.

One can differentiate between three sources of variability: internal variability due to the character of the function itself, external variability due to factors of the work environment, and aggregated variability due to functional upstream-downstream couplings.

After it has been identified, the performance variability has to be characterised using different variability manifestations, the phenotypes. There are two solutions: the simple one, which is efficient but not so thorough and considers only two phenotypes, i.e. timing and precision, and the elaborate approach that is not efficient but thorough and includes multiple phenotypes, i.e. speed, distance, sequence, object, force, duration, direction and timing. If we take a closer look at the simple solution, the output in terms of timing can occur too early, on time, too late or not at all. In terms of precision, the output can be precise, acceptable or imprecise (Hollnagel, 2017, p. 69-73).

3.1.5. Step 3: Aggregation of variability

However, it is not enough to simply know the variability of individual functions in isolation. In fact, the combination effects of several function variabilities has to be understood to know where functional resonance emerges. This is done by using the concept of aggregated variability to define upstream-downstream couplings. The variability can be caused by couplings of upstream functions, when the output used as input, precondition, resource, control or time is variable and thus affects the variability of downstream functions.

This impact can have three different expressions: variability is likely to increase (amplifying effect), variability is likely to decrease (damping effect) and variability is likely to stay unchanged (no effect). Table 1 illustrates this variability propagation for the combination of the variability phenotypes timing and precision and the five functional aspects.

3.1.6. Step 4: Management of variability

The final and last step includes the monitoring and management of the performance variability that was identified in the previous steps. The performance variability can lead to positive and negative effects. Positive effects should therefore be amplified by facilitating their occurrence without losing control, and negative effects should be dampened through elimination and prevention, though care should be taken not to eliminate the variability completely, since variability is essential for the safety and performance of the system (Hollnagel, 2017, p. 89).

FRAM only can offer pointers as to where to look but give no precise solutions. Thus, once the critical aspects and weaknesses have been identified, proper solutions have to be found and performance indicators should be established to monitor processes and developments to regulate the activities in a system. Therefore, variability can be dampened to a level where no unwanted outcomes arise – instead only desirable outcomes occur – and a safe and efficient working of the system is ensured.

Table 1
Upstream/downstream propagation of variability (Patriarca et al., 2018).

Upstream output variability		Input	Precondition	Resource	Control	Time
Timing Variability of Output	Too early	Amplifying/No Effect	Amplifying	No Effect/Damping	Amplifying	Amplifying
	On time	Damping	Damping	Damping	Damping	Damping
	Too late	Amplifying	Amplifying	Amplifying	Amplifying	Amplifying
	Not at all	Amplifying	Amplifying	Amplifying	Amplifying	Amplifying
Precision Variability of Output	Imprecise	Amplifying	Amplifying	Amplifying	Amplifying	Amplifying
	Acceptable	No Effect	No Effect	No Effect	No Effect	No Effect
	Precise	Damping	Damping	Damping	Damping	Damping

3.2. Case study: Overtaking manoeuvre on rural road

In this subsection, FRAM is applied to a concrete scenario in road traffic. First of all, it is important to define a reasonable scenario before executing the individual steps of FRAM. This process describes the underlying methodology and finally the results of each step.

3.2.1. Selection and description of scenario

The scenario is generally intended to represent a potential for automation, i.e. an accident black spot, and to provide sufficient material for a FRAM analysis and its evaluation. It should be noted, that the presented statistics for Germany below only serves as an example for the subsequent explanations, arguments and especially the scenario to be analysed. Thus, the scenario and the resulting FRAM-model can also be relevant for the settings of other countries.

Generally, the scenario has to fulfil the following requirements:

- the chosen accident location and cause of the accident should have a high number of people killed in traffic accidents
- the selected cause of the accident should consist of several different elementary driving tasks that serve the applicability or usefulness of the FRAM method
- the chosen cause of the accident has to ensure the interaction of the ego vehicle (own vehicle from the perspective of the driver) with at least one other road user

In order to meet the first criterion, a driving scenario has to be selected that takes place on rural roads. 58% of all fatal accidents in 2015 in Germany occurred on rural roads. The proportions for urban roads and highways are 30% and 12%, respectively (Destatis, 2016). Next, the contribution of the cause of the accident to the consequences of a fatal accident plays a role. On rural roads, the majority of accidents are driving accidents (38%) and accidents in longitudinal traffic (24%).

A driving accident is defined as a loss of control over the vehicle without any contribution from other road users. However, one result of uncontrolled vehicle movements may be a collision with other road users. Accidents in longitudinal traffic are conflicts between road users moving in the same or opposite direction (Destatis, 2017). These accidents are not related to a turn. That two conflict situations account for around three quarters (74%) of the fatalities in rural road accidents (Heinrich et al., 2010). In this respect, the main reasons are excessive speed, incorrect road use and poor overtaking opportunities (Destatis, 2016).

Overall, overtaking is the cause of a significant share of the accident types driving accident and accident in longitudinal traffic.

In addition, the overtaking task involves several different subtasks such as swerve, adjust the speed, merge, etc., so the second criterion is met. Meanwhile, the ego vehicle is in interaction with at least one other road user whose driving behaviour must be continuously considered by the driver of the ego vehicle, thus the third and last requirement is fulfilled. For these reasons, overtaking on a rural road was chosen as an adequate driving scenario for the investigation of this work.

In the scenario, the ego vehicle is on a long, flat straight section of a country road with one lane for each direction and a slower vehicle is driving in front. The rear and oncoming traffic is relatively far away and represents no immediate hazard. The aim of the ego vehicle is to overtake the vehicle in front. The weather is sunny, the road is in perfect condition, overtaking is permitted and no obstructions exist. The ego vehicle is driven once by a human driver and once by an automated system (SAE-level 4) according to SAE J3016 (2014) without any C2X. The vehicle in front is always driven by a human driver in both cases. Overall, this scenario represents a simple overtaking manoeuvre.

To get a better overview, the scenario can be divided into five segments (see Fig. 7): follow a vehicle in front, swerve into the oncoming lane, pass the leading vehicle, merge back into the starting lane and get in lane again. The entire scenario is a temporal sequence of actions (edges) and scenes (nodes) according to Ulbrich et al. (2015).

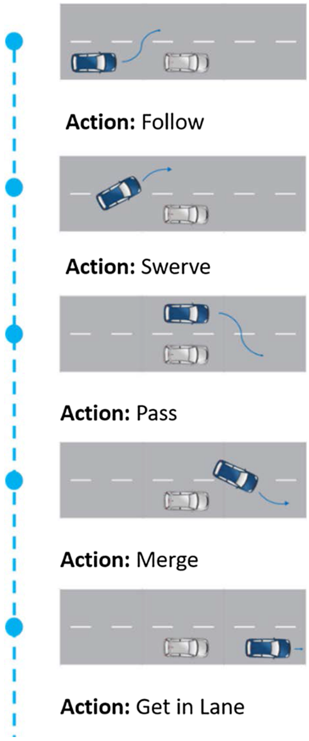


Fig. 7. The scenario “overtaking on a rural road” (dashed in blue) as a temporal sequence of actions/events (edges) and scenes (nodes) according to Ulbrich et al. (2015). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2.2. Applying step 1: Road system functions related to an overtaking manoeuvre

Initially, the situational analysis of the behavioural requirements of driving tasks (SAFE) in Fastenmeier & Gstalter (2007) was applied as a fundamental basis to identify and define the functions for the driving tasks later in FRAM. SAFE is a procedure for driving task analysis and driver requirement assessment. According the SAFE method, first the scenario under investigation is initially precisely defined and classified. Basic driving tasks are then derived and further subtasks are defined and segmented in time and space. These subtasks are further analysed based on a model of human information processing to compile all of the requirements that have to be fulfilled by the driver to correctly cope with the given task. Finally, these requirements are transformed into functions for FRAM.

In a second step, the functional decomposition of the road system (Kuzminski et al., 1995) was used to define further functions for the human driver that are missing if only SAFE is used, and especially to derive functions that relate to the vehicle or environment, which are in most cases background functions that are relevant for foreground functions. The use of the functional definition of the road system should ensure that as many functions as possible are identified and defined, or at least the most important ones.

After the two first steps, the functions and their couplings were integrated iteratively into a model using FMV, as shown in Figs. 8, 9 and 10. A total of three FRAM models had to be created, to keep an overview and facilitate the comprehension of the model. It should be noted that, basically, the three figures must be seen like three fundamental FRAM models for the overtaking manoeuvre and in the future different instantiations can be derived based on this to investigate several issues, for example various conditions of the environment by the weather or traffic density. So, they are rather toy-models and not different instantiations, even if, according to the respective terminology, this is an improper use of the wordings of model and instantiation. Fig. 8 shows the first FRAM model for the segments one and two, Fig. 9 the second FRAM model for segments three and four, and finally Fig. 10 presents the third FRAM

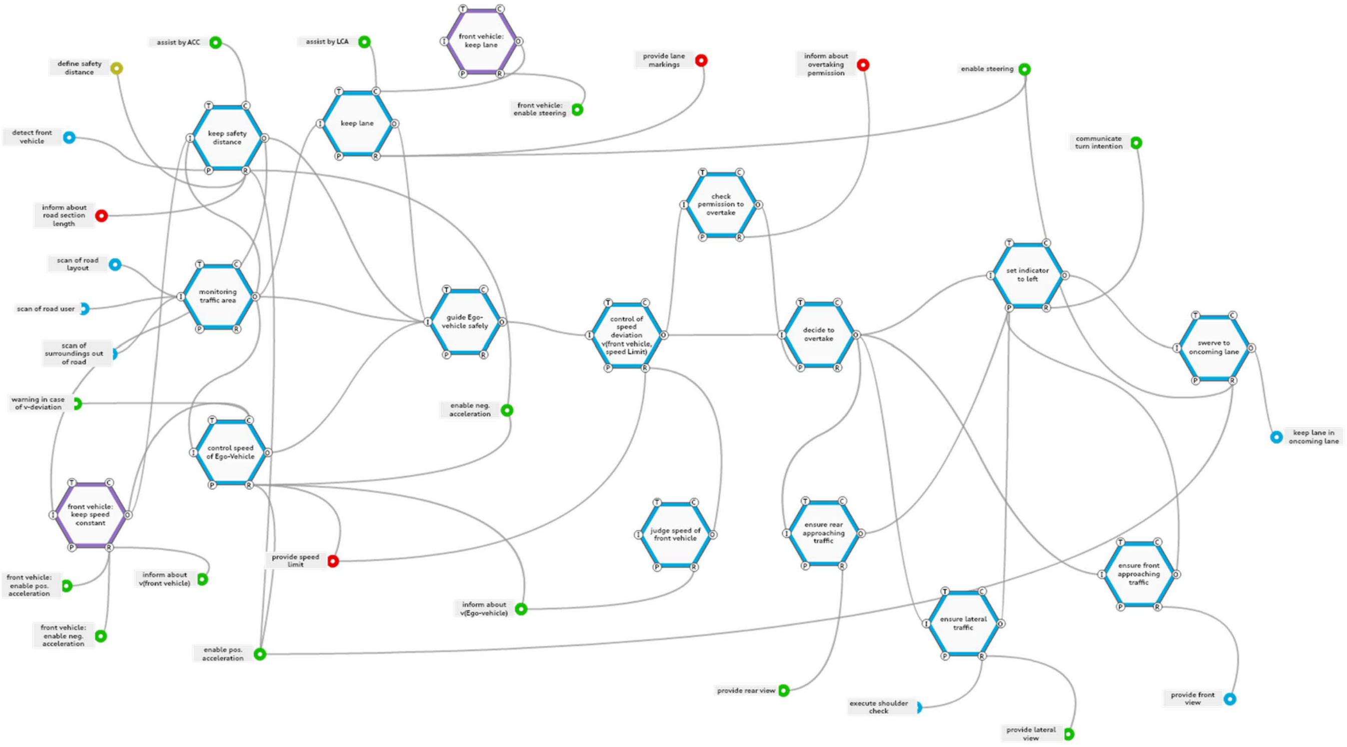

assist by ACC
assist by LCA
front vehicle: keep lane
provide lane markings
inform about overtaking permission
enable steering
define safety distance
keep safety distance
keep lane
front vehicle: enable steering
detect front vehicle
communicate turn intention
inform about road section length
check permission to overtake
scan of road layout
set indicator to left
monitoring traffic area
scan of road user
guide Ego-vehicle safely
control of speed deviation v(front vehicle, speed Limit)
decide to overtake
swerve to oncoming lane
scan of surroundings out of road
warning in case of v-deviation
enable neg. acceleration
keep lane in oncoming lane
control speed of Ego-Vehicle
front vehicle: keep speed constant
judge speed of front vehicle
ensure rear approaching traffic
ensure front approaching traffic
provide speed limit
front vehicle: enable pos. acceleration
inform about v(front vehicle)
inform about v(Ego-vehicle)
ensure lateral traffic
front vehicle: enable neg. acceleration
enable pos. acceleration
provide rear view
execute shoulder check
provide front view
provide lateral view

Fig. 8. FRAM model 1 for following and swerving in the overtaking manoeuvre scenario (segment 1 and 2). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

model for segment five. The hexagons indicate foreground functions and the rectangles stand for background functions. The different colours specify the various agents that perform the functions: human driver or automation driving the ego vehicle in blue, technology features of the vehicle in green, characteristics of the infrastructure in red, information by the policy in yellow and actions of the human driver of the leading vehicle in purple. In order to ensure that the three models are connected, and thus an entire and not a decomposed model will be created, the last foreground function of the preceding model is the first background function of the subsequent model and the first foreground function of the subsequent model is the last background function of the preceding model in each case as regards the temporal sequence. It should be noted that the

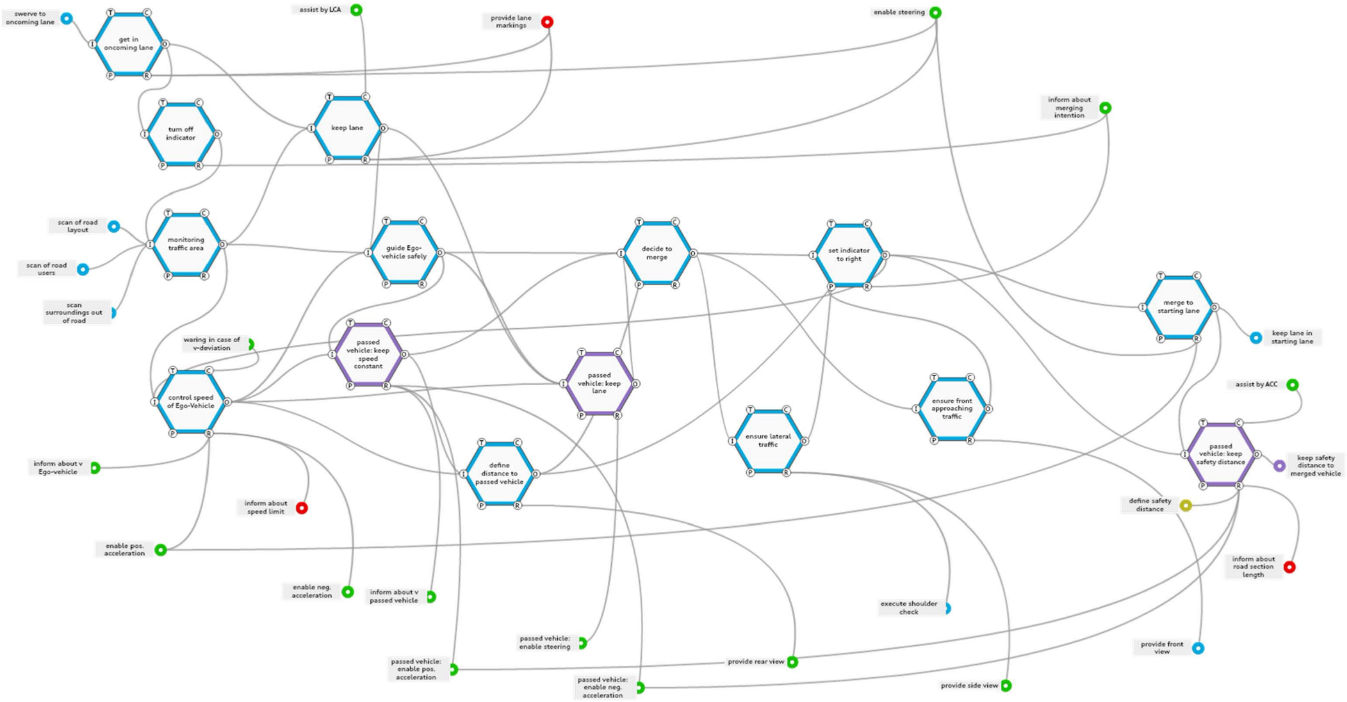

swerve to oncoming lane
get in oncoming lane
assist by LCA
provide lane markings
enable steering
inform about merging intention
turn off indicator
keep lane
scan of road layout
monitoring traffic area
guide Ego-vehicle safely
decide to merge
set indicator to right
scan of road users
merge to starting lane
keep lane in starting lane
scan surroundings out of road
waring in case of v-deviation
passed vehicle: keep speed constant
passed vehicle: keep lane
assist by ACC
control speed of Ego-Vehicle
ensure front approaching traffic
ensure lateral traffic
define distance to passed vehicle
passed vehicle: keep safety distance
keep safety distance to merged vehicle
inform about v Ego-vehicle
define safety distance
inform about speed limit
enable pos. acceleration
inform about road section length
enable neg. acceleration
inform about v passed vehicle
execute shoulder check
passed vehicle: enable steering
provide front view
passed vehicle: enable pos. acceleration
provide rear view
passed vehicle: enable neg. acceleration
provide side view

Fig. 9. FRAM model 2 for passing and merging in the overtaking manoeuvre scenario (segment 3 and 4). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

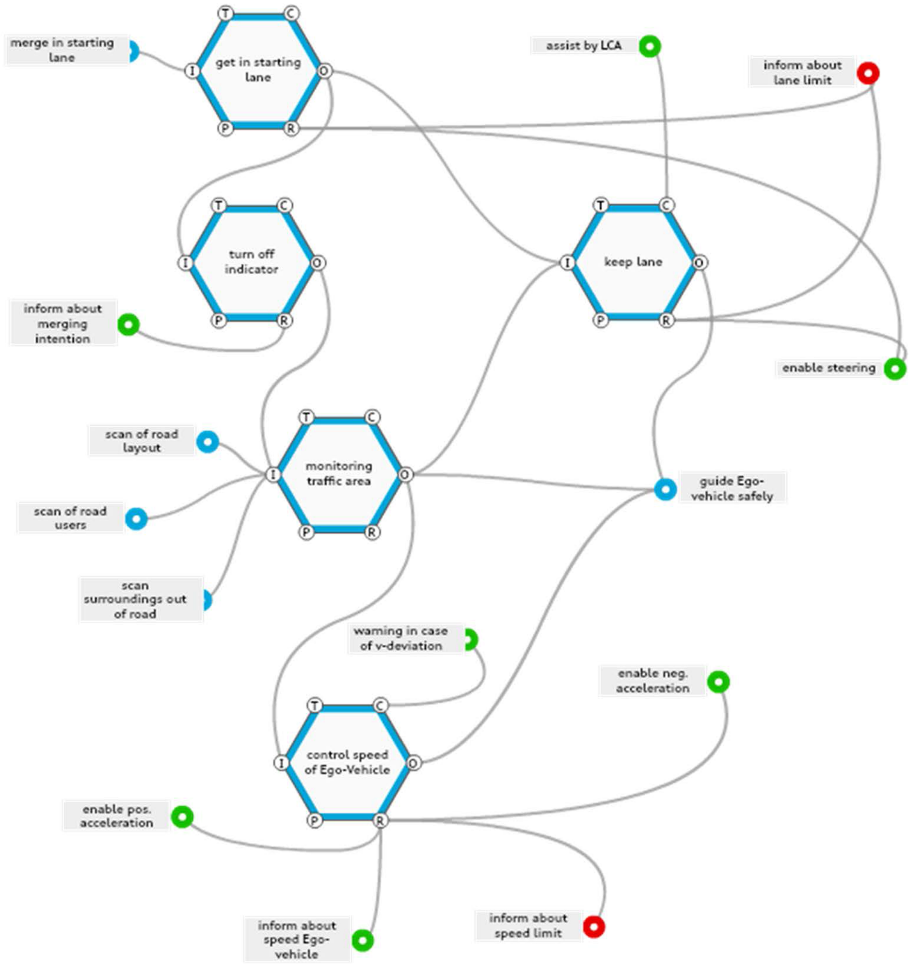


Fig. 10. FRAM model 3 for getting into starting lane again in the overtaking manoeuvre scenario (segment 5). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

models are the same for the human driver or the automation due to the assumption that there is no change in the functions of the system that have to be accomplished by the human driver or the automation. The difference between the two agents is only the variable performance of each function.

Tables 2 and 3 list the foreground- and background functions involved in the analysis, detailing the agent performing them. There are a total of 26 foreground and 27 background functions, whereby some functions are repeated during the overall overtaking manoeuvre.

3.2.3. Applying step 2: Identification of performance variability for the driving task

In the second step according to FRAM, the actual variability of each foreground function is characterised individually based on internal and external variability sources following Hollnagel's (2017) simple solution, i.e. in terms of timing and precision. The authors assigned the manifestation of variability subjectively based on their experience in the field as well as knowledge about the current state of driving assistance systems and automated driving systems. Consequently, the data was not obtained empirically, but it did not have to be because this work is not aimed at content-related results but at methodological results regarding the applicability of FRAM. It does not matter if the variability is realistic. Table 4 shows the assigned manifestations of variability for timing and precision in both cases, driving by a human and automation. It is assumed that the variability of the functions related to the vehicle in front or passed vehicle is the same in both cases.

In addition, the semi-quantitative approach according to Patriarca et al. (2017b) was applied to enhance the graphical and qualitative approach of the traditional use of FRAM which is difficult to read for highly complex systems. Thus, a numerical score was assigned to each performance variability state in a first step (see Table 5). The higher the score, the more variable the output. The variability of the upstream output j, OV_j is the product of these two scores (1):

$$OV_j = V_j^T \bullet V_j^P \tag{1}$$

where

V_j^T represents the upstream output j score in terms of timing
V_j^P represents the upstream output j score in terms of precision

3.2.4. Applying step 3: Aggregating the performance variability quantitatively for the road system

Once assigned the variability score for the upstream output, the coupling variability (CV) of the upstream output j and the downstream function i (2) as well as associated variability propagation factors $a_{ij}{}^T$ and $a_{ij}{}^P$ have to be specified according to Patriarca et al. (2017b) (3):

$$CV_{ij} = OV_j \cdot a_{ij}{}^T \cdot a_{ij}{}^P \tag{2}$$

where

$a_{ij}{}^T$ represents the propagation factor for the upstream output j and the downstream function i in terms of timing
$a_{ij}{}^P$ represents the propagation factor for the upstream output j and the downstream function i in terms of precision

Table 2
List of foreground functions related to the performing agent.

Foreground function	Agent
to keep safety distance	human driver/automation (ego vehicle)
to keep lane	human driver/automation (ego vehicle)
to monitor traffic area	human driver/automation (ego vehicle)
to control speed of ego vehicle	human driver/automation (ego vehicle)
to guide ego vehicle safely	human driver/automation (ego vehicle)
front vehicle: to keep lane	human driver (front vehicle)
front vehicle: to keep speed constant	human driver (front vehicle)
to control speed deviation between velocity of front vehicle and speed limit	human driver/automation (ego vehicle)
to check permission to overtake	human driver/automation (ego vehicle)
to judge speed of front vehicle	human driver/automation (ego vehicle)
to decide to overtake	human driver/automation (ego vehicle)
to ensure rear approaching traffic	human driver/automation (ego vehicle)
to ensure lateral traffic	human driver/automation (ego vehicle)
to ensure front approaching traffic	human driver/automation (ego vehicle)
to set indicator to the left	human driver/automation (ego vehicle)
to swerve into oncoming lane	human driver/automation (ego vehicle)
to get into oncoming lane	human driver/automation (ego vehicle)
to turn off indicator	human driver/automation (ego vehicle)
passed vehicle: to keep speed constant	human driver (front vehicle)
passed vehicle: to keep in lane	human driver (front vehicle)
to define distance to passed vehicle	human driver/automation (ego vehicle)
to decide to merge	human driver/automation (ego vehicle)
to set indicator the right	human driver/automation (ego vehicle)
to merge into starting lane	human driver/automation (ego vehicle)
passed vehicle: to keep safety distance	human driver (front vehicle)
to get into starting lane	human driver/automation (ego vehicle)

Table 3
List of background functions related to the performing agent.

Background function	Agent
to define safety distance	policy
to detect front vehicle	human driver/automation (ego vehicle)
to inform about road section length	infrastructure
to scan road layout	human driver/automation (ego vehicle)
to scan road user	human driver/automation (ego vehicle)
to scan surroundings of the road	human driver/automation (ego vehicle)
to warn in case of illegal speed deviation	technology
front vehicle: to enable positive acceleration	human driver (front vehicle)
front vehicle: to enable negative acceleration	human driver (front vehicle)
to inform about speed of front vehicle	technology
to enable pos. acceleration	technology
to enable negative acceleration	technology
to provide speed limit	infrastructure
to inform about speed of ego vehicle	technology
to provide rear view	technology
to execute shoulder check	human driver/automation (ego vehicle)
to provide lateral view	technology
to provide front view	technology
to communicate turn intention	technology
to enable steering	technology
to inform about overtaking permission	infrastructure
to provide lane markings	infrastructure
front vehicle: to enable steering	technology
to assist by LCA	technology
to assist by ACC	technology
to keep safety distance to merged vehicle	human driver (front vehicle)
to inform about merging intention	technology

Note that $a_{ij}{}^{T}$ or $a_{ij}{}^{P}$ may assume the following values:

$$\begin{array}{ll} 2 & \text{if the upstream output has an amplifying effect on the downstream function} \\ 1 & \text{if the upstream output has no effect on the downstream function} \\ 0.5 & \text{if the upstream output has a damping effect on the downstream function} \end{array} \tag{3}$$

The specification of the propagation factor is based on Table 1.

The downlink (*DL*) and uplink (*UL*) coupling variability of one foreground function (downlink functional coupling variability, *DLFCV* and uplink functional coupling variability *ULFCV*) should be calculated in the next step. The *DLFCV* is used to understand the implications of the coupling variabilities of one entire upstream function *j* to associated downstream functions *i* and the *ULFCV* is used to comprehend the impact of the variability of a downstream function *i* through its incoming coupling variabilities of upstream functions *j*. The calculation formula for *DLFCV* and *ULFCV* can be seen in (4) and (5), respectively:

$$DLFCV_{ij} = \sum_{i=1}^{j} CV_{ij} \tag{4}$$

$$ULFCV_{ji} = \sum_{j=1}^{i} CV_{ij} \tag{5}$$

Additionally, the number (*N*) of downlinks of an upstream function *j* ($N_{DL}{}^{j}$) and the number of uplinks of a downstream function *i* ($N_{UL}{}^{i}$) have to be determined. This allows the number of links of an upstream function to downstream functions or vice versa to be specified. $N_{DL}{}^{j}$ is the sum of downlinks of an upstream function (6) and $N_{UL}{}^{i}$ is the sum of uplinks of a downstream function (7):

Table 4
Variability manifestations for each function in a comparison of the human driver and automation.

Foreground function	Timing of human driver	Precision of human driver	Timing of automation	Precision of automation
to keep safety distance	too late	acceptable	on time	precise
to keep lane	too late	acceptable	on time	precise
to monitor traffic area	on time	precise	on time	imprecise
to control speed of ego vehicle	too late	acceptable	on time	precise
to guide ego vehicle safely	too late	acceptable	on time	acceptable
front vehicle: to keep lane	on time	acceptable	–	–
front vehicle: to keep speed constant	on time	acceptable	–	–
to control speed deviation between velocity of front vehicle and speed limit	on time	precise	on time	precise
to check permission to overtake	on time	precise	on time	acceptable
to judge speed of front vehicle	on time	imprecise	on time	precise
to decide to overtake	on time	precise	Too late	acceptable
to ensure rear approaching traffic	on time	precise	on time	acceptable
to ensure lateral traffic	on time	precise	on time	acceptable
to ensure front approaching traffic	on time	precise	on time	acceptable
to set indicator to the left	too late	precise	on time	precise
to swerve into oncoming lane	too early	acceptable	on time	precise
to get into oncoming lane	too late	acceptable	on time	precise
to turn off indicator	on time	precise	on time	precise
passed vehicle: to keep speed constant	too late	acceptable	–	–
passed vehicle: to keep lane	too late	acceptable	–	–
to define distance to passed vehicle	too late	acceptable	on time	precise
to decide to merge	on time	precise	on time	acceptable
to set indicator the right	too late	precise	on time	precise
to merge into starting lane	too early	acceptable	on time	precise
passed vehicle: to keep safety distance	too late	acceptable	–	–
to get into starting lane	too late	acceptable	on time	precise

Table 5
Assignment of numerical values to the linguistic description of variability manifestation of the phenotypes timing and precision.

Variability phenotype	Variability manifestation	V_j^T or V_j^P
Timing	too early	2
	on time	1
	too late	4
	not at all	5
Precision	imprecise	5
	acceptable	3
	precise	1

$$N_{DL}{}^{j} = \sum_{i=1}^{j} DL_{ij} \tag{6}$$

$$N_{UL}{}^{i} = \sum_{j=1}^{i} UL_{ji} \tag{7}$$

It should be mentioned that only the downlinks or uplinks between two foreground functions and not between two background functions or between a foreground and a background function are counted.

In the final step, a global system variability (*GSV*) can be calculated that is the product of n functions within the whole system consisting of the multiplication of each associated *DLFCV* with the respective number of downlinks (8):

$$GSV = \prod_{j=1}^{n} DLFCV_{ij} * N_{DL,j} \tag{8}$$

The number of downlinks functions as a kind of weighting factor related to the impact of the associated *DLFCV*. The variability is quantified with the help of the software myFRAM (Patriarca et al., 2017c), which was developed in Visual Basic for Applications (VBA) and interfaced with Microsoft Excel and FMV. The main purpose of myFRAM is to develop and explore a FRAM model in a systematic way and to enhance FRAM for further analyses, such as the semi-quantitative approach in Patriarca et al. (2017b) described above.

The quantified data of the formulas can be visualised in diagrams that are created in Excel. Two diagrams can be seen in Figs. 11 and 12 as an example of the functional coupling variability outputs with regard to the first FRAM model (in Fig. 8). Similarly, this is also possible for the second and third FRAM model. Fig. 11 shows the functional coupling variabilities and number of links for the human driver and Fig. 12 for the automation. The number of downlinks (blue columns) and uplinks (orange columns) is shown on the left y-axis. The *DLFCV* (transparent area in blue) and *ULFCV* (transparent area in orange) are presented on the right y-axis and the different functions are on the x-axis. It should be noted that the diagrams represent a static state of the model and must be calculated repeatedly step by step for dynamic progressions, which is more realistic. Furthermore, the number of links is the same for the human driver and automation due to the assumption that there is no change in the functions of the system that have to be accomplished by the human driver or the automation as described above in Section 3.2.2. The only difference between the two agents that can be seen is in the functional coupling variabilities. These figures can be used to argue that, on the one hand, the higher the *DLFCV* and the greater the number of downlinks of one upstream function, the more critical the variable output of this function will be, and on the other hand, the higher the *ULFCV* and the greater the number of uplinks of one downstream function, the more likely it is that the output of this function will be highly variable, thus more critical. Or another interpretation example could be if the *ULFCV* is high and the *DLFCV* is low of the same function then this could mean that the incoming variability is tolerated and has no significant impact on the output variability. Overall, these figures should be used to gain a better and quick idea of where the critical functions in the system lie and then to delve deeper into understanding why and how this criticality occurs using the graphical illustration in FMV. Finally, countermeasures have to be considered to reduce this criticality by dampening the variability.

The *GSV* is illustrated in Fig. 13, where the *GSV* is defined for each single FRAM model (F1, F2, F3) and in a combination of the first two FRAM models (F1*F2) or all three (F1*F2*F3). Additionally, this figure shows a comparison between the human driver (blue) and the

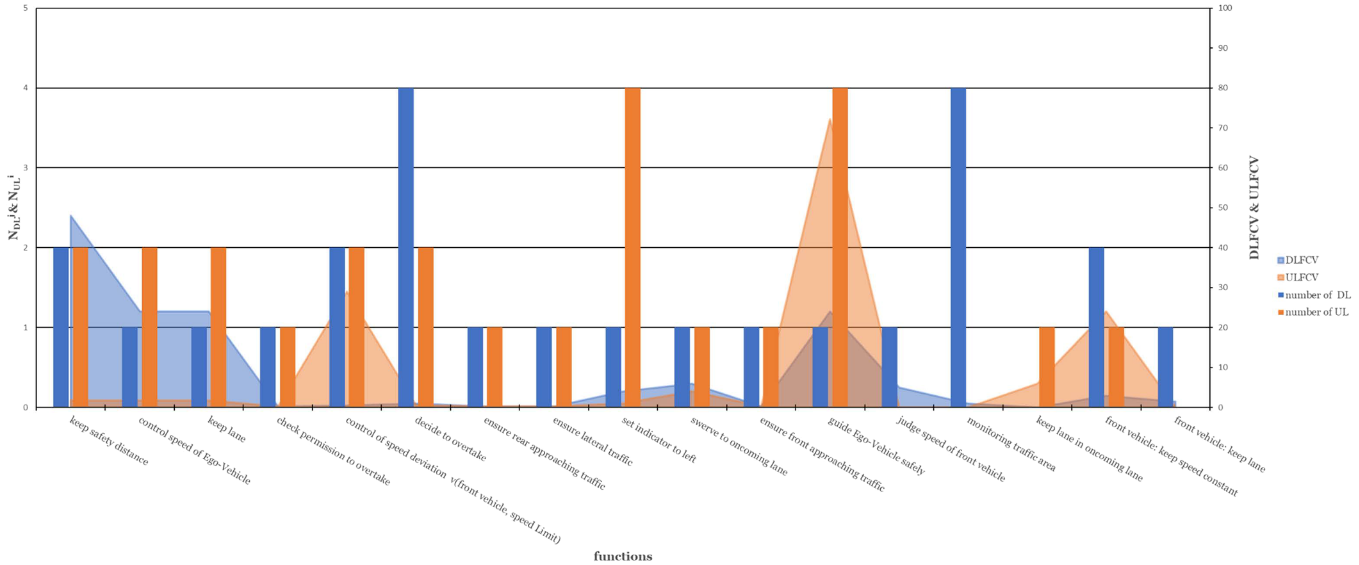


Fig. 11. FRAM model 1 for the human driver - Functional coupling variabilities and number of links. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

automation (orange). The value of *GSV* is outlined on a logarithmic scale on the x-axis. On the whole, this diagram could be used to illustrate the overall variability within the system for one instantiation or scenario in a comparison of two different system designs. By means of this illustration, it can be concluded that the higher the variability, the less safe the system will be.

3.2.5. *Applying step 4: Monitoring and managing the variability in the road context*

Once created, the graphical FRAM models and the different quantified outputs have to be interpreted and results derived from them. If we take a closer look at Figs. 11 and 12, we can see that the *DLFCV* increases moderately for the function *to keep safety distance* with two downlinks and slightly for the functions *to control speed of ego vehicle, to keep in lane* and *to guide ego vehicle safely,* each with one downlink, for the human driver. On the other hand, the *DLFCV* increases significantly for the function *to decide to overtake* and slightly for the function *to monitor traffic area,* each with four downlinks, for the automation. This shows that the human driver has more trouble coping with driving tasks related to stabilisation of the vehicle on the road, but that these functions have a low potential impact on other downstream functions due to the small number of downlinks. In contrast, the automation is riddled with decision-making tasks for overtaking or sensing tasks such as monitoring the traffic, and these functions have a great potential effect on other downstream functions due to the relatively high number of downlinks.

Thus, one interpretation could be that the effects of the automation are more critical than those of the human driver and that full automation is unreasonable. Alternatively, you could simply follow an ADAS approach and say that the functions with increased *DLFCV* for the human driver should be automated and the functions with increased *DLFCV* for the automation should continue to be performed by human drivers.

Further on, it can be seen that the *ULFCV* is slightly higher for the functions *to control speed deviation between ego- and vehicle in front* and *to keep speed constant by vehicle in front,* with two or one uplinks, and much

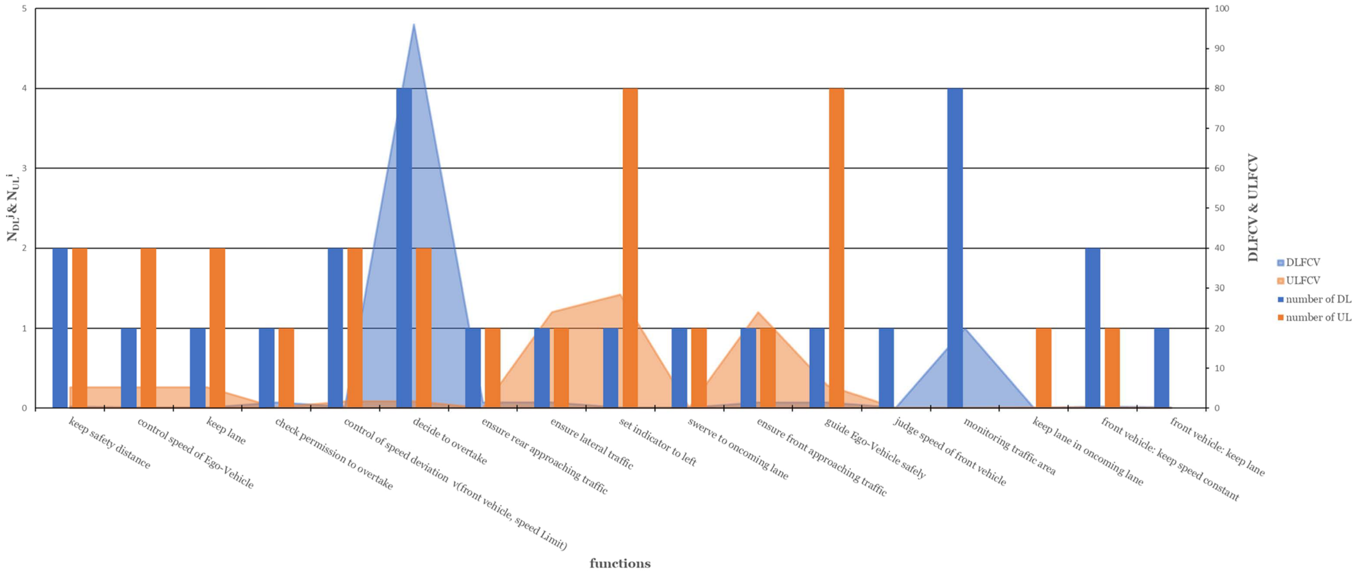


Fig. 12. FRAM model 1 for the automation - Functional coupling variabilities and number of links. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

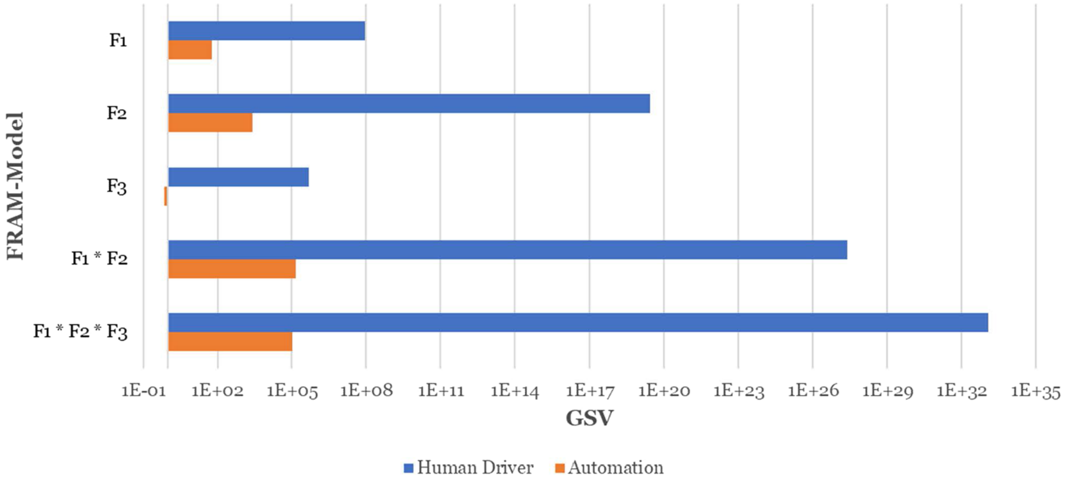


Fig. 13. Global system variability (*GSV*) for each single FRAM model and as a combination of these in a comparison between the human driver and the automation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

higher for the function *to guide ego vehicle safely,* with four uplinks, for the human driver. In return, the *ULFCV* is slightly higher for the functions *to ensure lateral traffic, to set indicator to left* and *to ensure front approaching traffic,* with one, four or one uplinks, for the automation.

Therefore, one interpretation may be that incoming variabilities are collectively more critical for the human driver than for automation, for example to guide the ego vehicle safely. This is why the upstream functions of this function, namely *to keep safety distance, to keep in lane, to monitor traffic area* and *to control speed of ego vehicle,* should be automated. This could be regarded as a confirmation of the previous interpretation of results from *DLFCV*.

After the functional coupling variabilities and number of links have been interpreted, providing a good overview of where to look in more detail, the critical functions and related aspects need to be differentiated and explored in the graphical model using FMV. Let us take the function *to guide the ego vehicle safely,* which is referred to in the following text as a function in the focus (FiF), as an example. The mechanisms for the comparison of the human driver and the automation are illustrated in the Figs. 14 and 15, which show an excerpt from the FRAM model 1 with a focus on the dependencies between the upstream and downstream functions related to the function FiF. These couplings are highlighted in purple. The value for *ULFCV* is shown above the hexagon of the FiF and the value for *DLFCV* below this. The respective value of *CV* is displayed on each purple line. This information visualises the composition of the *ULFCV* and you can see where so much variability comes from. Furthermore, each hexagon has an upper and lower coloured line according to the new feature of FMV Pro 2.0 to trace the variability in a simple form. The upper line and lower line represent the variability score for timing and precision, respectively. The resulting colours on the functions are not fixed like the assigned variability manifestations in Table 4. They are in fact rendered as continuous effects of upstream and downstream couplings and the scale varies between blue and red (with green as the mid or neutral point). Specifically, the colour coding is as follows: too early (orange), on time (blue), too late (orange), not at all (red), precise (blue), acceptable (green) and imprecise (red).

We can see that the FiF has four potential sources of incoming variability, whereby, the FiF has three main sources for the human driver and only one major origin for the automation. In addition, the main variability source of the automation is not included in the three roots of variability for the human driver. So, there is a difference. The human driver is more variable in terms of stabilisation tasks such as keeping in lane or the safety distance, whereas the automation is more variable in sensing tasks like monitoring the road to identify the road layout or other road users. One reason for FiF's higher *DLFCV* for the human driver compared to the automation may be that FiF’s output is affected by three increased *CV*’s and not just one which is less variable.

It can also be seen that the function *to monitor traffic area* also affects the other functions that directly influence the FiF, so that this function has a potentially high impact and is thus critical in terms of automation. This case is shown in Fig. 16 as “what happens if”. If we set the timing variability of the function *to monitor traffic area* from *on time* to *too late,* this leads to an increased timing variability in six other functions, highlighted in green in Fig. 16. Actually, this change indirectly affects the speed keeping of vehicle in front as there is a feedback loop between the ego vehicle keeping the safety distance that is directly influenced by monitoring the traffic area. Hence, this is a good example to show emergent effects in the road system. However, it is not obvious that the sensing task of the driver to monitor the traffic can affect the speed behaviour of the leading vehicle.

There are many more things to be analysed, but the previous interpretation examples should be sufficient to show the applicability of FRAM to road traffic.

Last but not least, the comparison of the *GSV* for the human driver and the automation in Fig. 13 can be interpreted as follows: the *GSV* related to the human driver is significantly higher than that related to the automation in both cases for every single FRAM model and for the models in common consideration. Thus, the overall scenario is more variable, i.e. unsafe, if the vehicle is driven by the human rather than by the automation. Consequently, this scenario, which is based on the predefined instantiation and associating assumptions should be automated. But we have to bear in mind two aspects that take a critical look at the *GSV*. First of all, a higher variability is per se not more unsafe. As already mentioned, a complex system needs variability to emerge safety, so the sources and reasons for this variability have to be differentiated. For example, at one point in time and space, this supposed negative variability seems to be reasonable and thus leads to a safe outcome, whereas the variability results in an accident at another point in time and space. Secondly, the calculation of *GSV* does not seem to fit the real system behaviour because there could be upstream functions that have a greater effect on downstream functions than others. In other words, some functions make a more critical contribution to accidents than others. This cannot be mapped solely by the product of the *DLFCV* and the number of downlinks. Therefore, an additional weighting factor has to be considered for each output of a downstream function and maybe specified empirically. Moreover, each downstream function has a certain robustness or tolerance factor towards the incoming variability. This too is not reflected in the current calculation of the *GSV*.

All in all, the consideration of the *GSV* is to be treated with caution and should be seen as a relative tendency rather than an absolute and valid fact. It is thus crucial to delve deeper into the model and

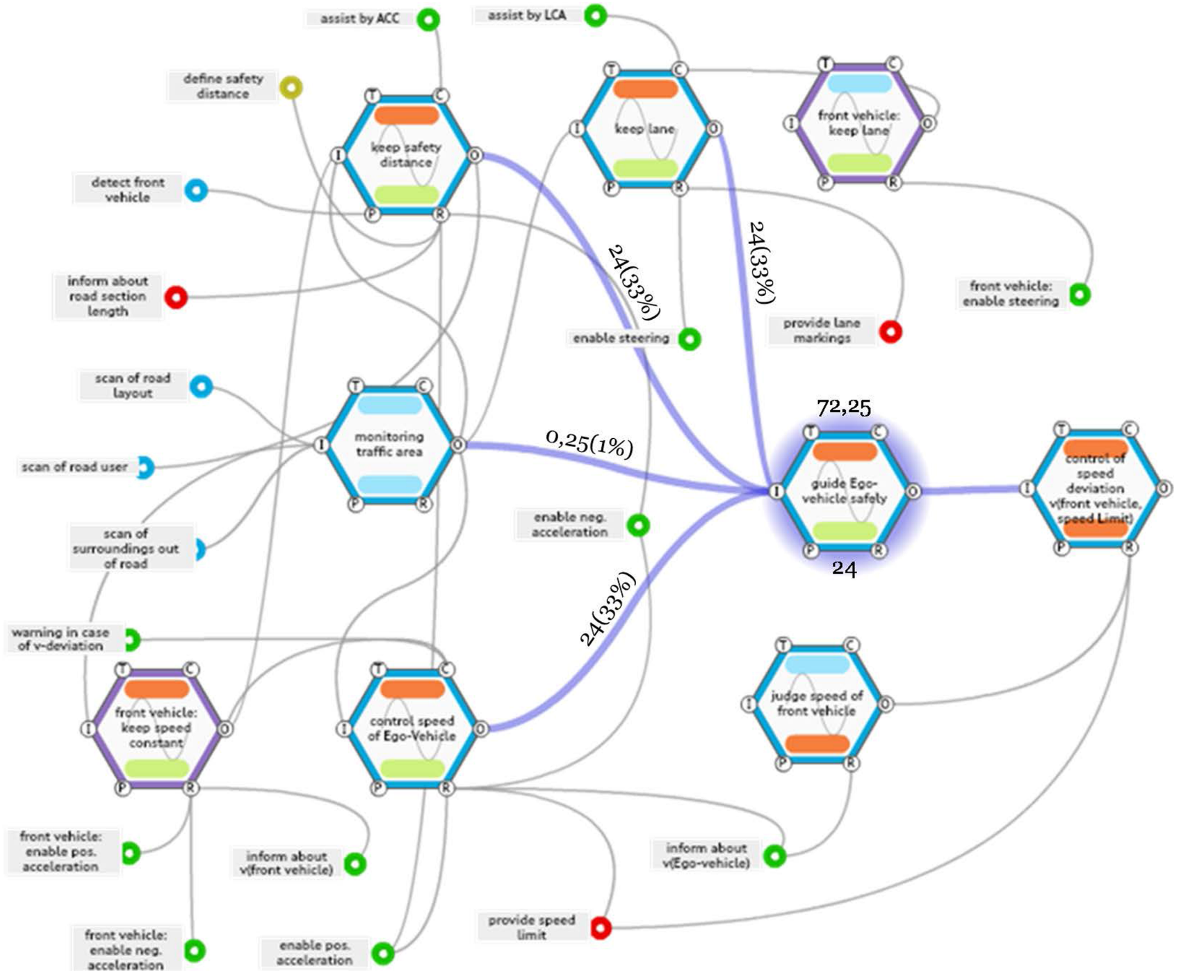

assist by ACC
assist by LCA
define safety distance
keep safety distance
keep lane
front vehicle: keep lane
detect front vehicle
inform about road section length
24(33%)
24(33%)
enable steering
provide lane markings
front vehicle: enable steering
scan of road layout
monitoring traffic area
72,25
0,25(1%)
scan of road user
guide Ego-vehicle safely
control of speed deviation v(front vehicle, speed Limit)
scan of surroundings out of road
enable neg. acceleration
24
24(33%)
warning in case of v-deviation
judge speed of front vehicle
front vehicle: keep speed constant
control speed of Ego-Vehicle
front vehicle: enable pos. acceleration
inform about v(front vehicle)
inform about v(Ego-vehicle)
provide speed limit
front vehicle: enable neg. acceleration
enable pos. acceleration

Fig. 14. One excerpt from the FRAM model 1 to show a critical path (purple) for the human driver related to the function *to guide the ego vehicle safely.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

understand the mechanisms and emergent and non-linear effects such as those shown above, and to not simply trust the global system variability.

4. Discussion

The aim of this paper, as mentioned in the introduction, is to identify or define a method that should be applied to specific and reasonable traffic scenarios so as to identify the contribution of the human driver to road traffic safety within these situations and to derive requirements for the automation and the potential of automation in these situations with its accompanying factors. For this purpose, the method has to differentiate between the mechanisms of road traffic and identify the interdependencies between each system element. Ultimately, design recommendations for automation as well as supporting material to reduce the validation effort should be provided using this method. Finally, the applicability of this method has to be evaluated in a case study.

First of all, the derivation of an adequate method was shown in great detail by a description of the development of safety thinking and risk assessment methods, the properties of the road traffic system, and last but not least, an explanation of why systemic methods, and in particular FRAM, should be used. The result was that FRAM has the highest potential to be the method we are searching for.

In the second part of the paper, FRAM was applied to an overtaking scenario to describe how this method can be used and to assess the suitability of FRAM for the underlying objective of this paper. Thus, the strengths as well as limitations of FRAM have to be discussed.

First, the strengths are explained. FRAM is very flexible to use since it is a method-sine-model. This also means that a FRAM model can be augmented or changed every time in terms of its granularity and has no limitations regarding modelling so that further users do not have to start from scratch. The "openness" of FRAM provides extensive opportunities for the combination of a FRAM analysis with various other tools and approaches (Patriarca et al. 2017a,b; Patriarca et al., 2018; Tian et al. 2016, among others), thus paving the way for an analysis of specific problems whilst maintaining an overall socio-technical system perspective (Ferreira and Cañas, 2019). Moreover, a generated model can be represented in a graphical form, so one does not simply see the input and output but rather a "map" between the input and output and how inputs are transferred into the specific outputs. Besides, practitioners have guidance material in form of the fundamental theory book of FRAM by Hollnagel (2017), a brief guide on how to use the FRAM (Hollnagel, 2018b) and a practical guideline or handbook (Hollnagel et al., 2014). In addition, the use of FRAM is supported by software, i.e. FMV and myFRAM, thus ensuring a kind of standardisation and systematic implementation. Besides, add-ons of the currently available software and also new software (EZ-FRAM) were announced at the FRAMily workshop 2019 in Malaga. The new software should support the simulation of how variability propagate through a FRAM model as well as considering the dynamics in a system or model. Additionally, to the pure and original qualitative approach of FRAM, a quantitative risk or safety assessment is also possible (Patriarca et al., 2017b) which is supported by myFRAM. Thus, the results of FRAM can be presented to a

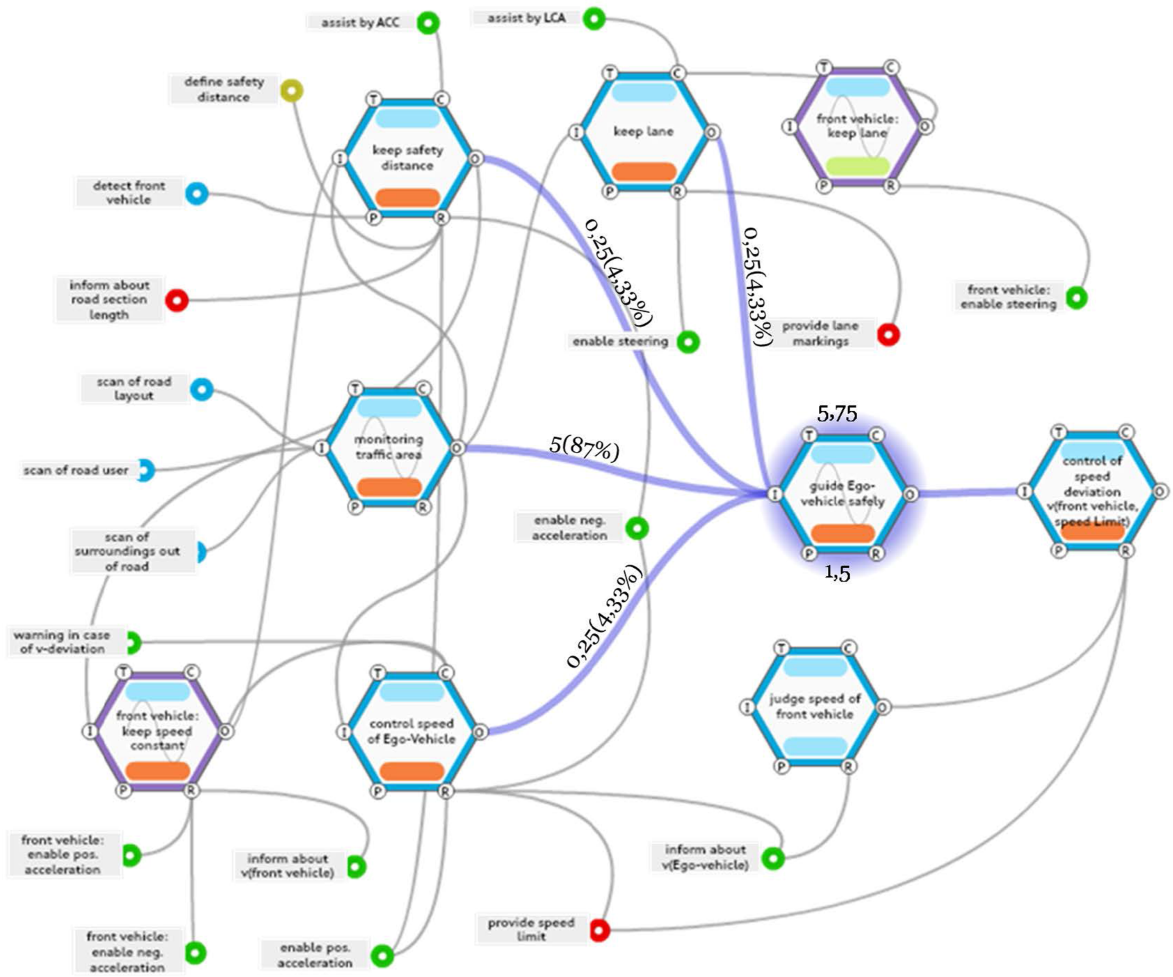

assist by ACC
assist by LCA
define safety distance
keep safety distance
keep lane
front vehicle: keep lane
detect front vehicle
0,25(4,33%)
0,25(4,33%)
inform about road section length
front vehicle: enable steering
enable steering
provide lane markings
scan of road layout
5,75
monitoring traffic area
5(87%)
scan of road user
guide Ego-vehicle safely
control of speed deviation v(front vehicle, speed Limit)
enable neg. acceleration
1,5
scan of surroundings out of road
0,25(4,33%)
warning in case of v-deviation
judge speed of front vehicle
front vehicle: keep speed constant
control speed of Ego-Vehicle
front vehicle: enable pos. acceleration
inform about v(front vehicle)
inform about v(Ego-vehicle)
provide speed limit
front vehicle: enable neg. acceleration
enable pos. acceleration

Fig. 15. One excerpt from the FRAM model 1 to show a critical path (purple) for the automation related to the function *to guide the ego vehicle safely*. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

wider audience more easily and the inevitable need of traditional risk approaches to describe risk and safety in terms of numerical values or statistical outputs is satisfied. Although, a quantification of FRAM is not mandatory or reasonable (Hollnagel, 2017, p. 94), it may be able to help with specific issues, in particular facilitating the interpretation of FRAM models related to large-scale complex systems such as the road system, as stated also by Ferreira and Cañas (2019). Finally, FRAM helps to identify the critical functions and their consequences for the entire system, and to visualise the mechanisms and emergence of events, specifically equi- and multifinality.

Next, the limitations are presented. FRAM is a very elaborate method and intense introductory work coupled with extensive domain and human factors knowledge (Hollnagel & Speziali, 2008) is needed before a FRAM analysis can begin. Actually, even with this relatively simple toy-model the FRAM model requires a lot of time resources of the researcher (approximately 40h). So, this will again increase significantly for more differentiated models and a methodological approach, whereby empirical data are collected. Similar observations were made by Adriaensen et al. (2019), while suggesting a limitation of the scope of a model to the essential questions under investigation to ensure a manageable model. But this worthwhile for a safety assessment of high-risk systems such as an automated vehicle. Furthermore, the illustration of a FRAM analysis in a graphical form for highly complex systems quickly becomes confusing due to its messy appearance. A sensible interpretation is therefore difficult, if not impossible, but a better understanding of the potential system dynamics is possible. Fortunately, this limitation can be eased by the approaches of Patriarca et al. (2017a,b) or Patriarca et al. (2018) and also through the use of myFRAM. The more critical aspects of limitation are the identification of system functions and their interdependencies as well as their variabilities. The current approach to identify functions and their variabilities is to study reports, procedures, design specifications, storytelling or to conduct field observations or interviews. Some practical guidance material exists in Hollnagel et al. (2014), but this is ultimately subject to a very strong subjective assessment of SME's. Regarding the objective of this paper and application of FRAM to assess the safety of automated driving, the identification of functions and their variability have to be improved in further research, which must include more objective and empirical measures. In addition, the validation of a FRAM model is impossible and calibration can only be achieved in the form of face validity (Bridges et al., 2018). Calibration means that a FRAM model is developed and evaluated iteratively by a group of SMEs' until every SME agrees with completeness and precision of the created model. Since the objective target of FRAM in this paper is to derive requirements for automated system design and to offer hints for its validation, but not to validate or approve the automated system based on the FRAM model per se (see also Fig. 2), this validation limitation is irrelevant. Nevertheless, for the future research it should be the ambition to calibrate a FRAM model by SME's to ensure the reliability of the results based on a FRAM analysis. Besides, this process should be enhanced by more objective, empirical and analytical approaches. Thereby, the transformation of a FRAM model into a Bayesian Belief Net (Slater, 2016) may be an appropriate approach. Also, against the background of the qualitative nature of the FRAM method the proposal

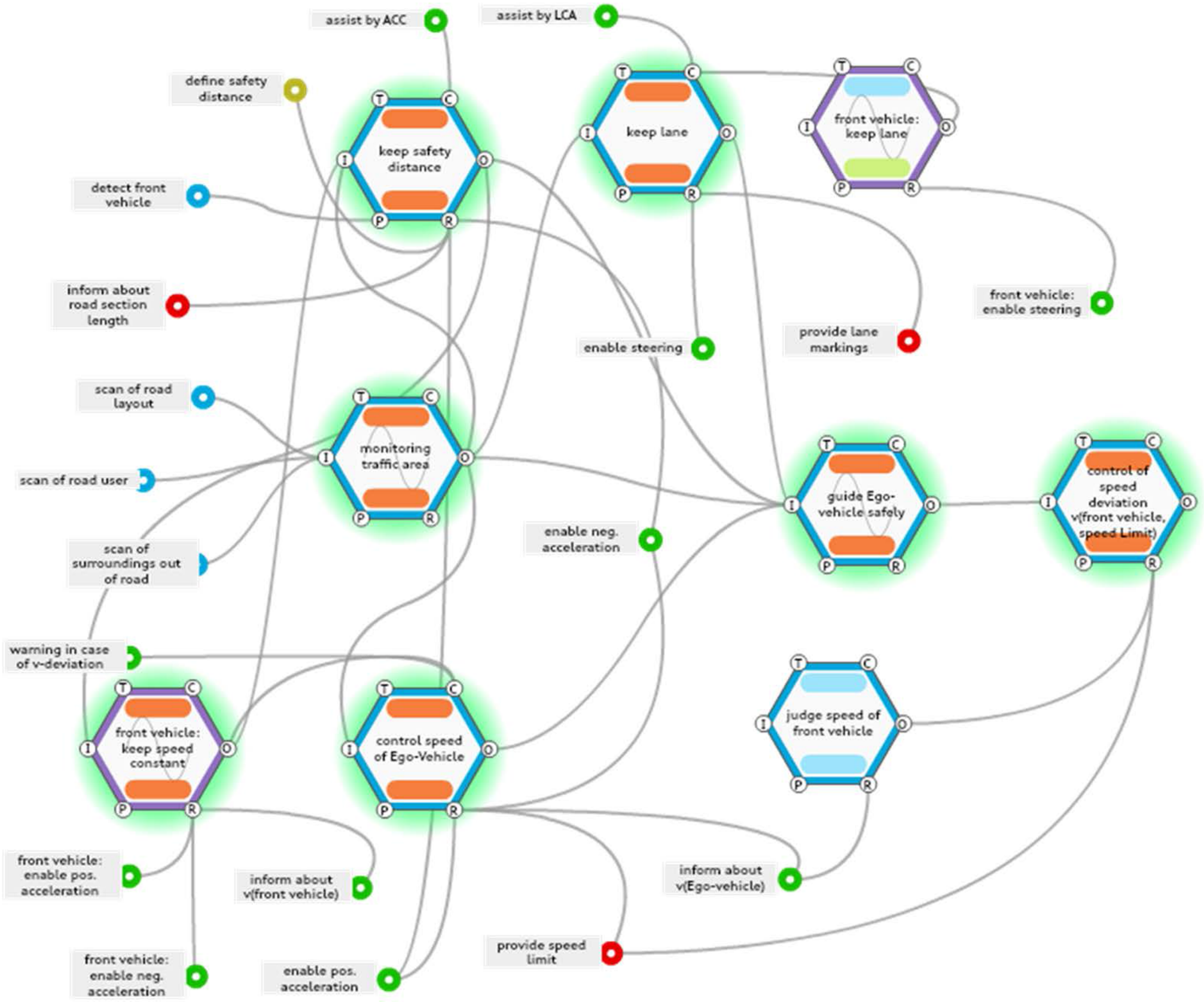

assist by ACC
assist by LCA
define safety distance
keep safety distance
keep lane
front vehicle: keep lane
detect front vehicle
inform about road section length
front vehicle: enable steering
provide lane markings
enable steering
scan of road layout
monitoring traffic area
scan of road user
guide Ego-vehicle safely
control of speed deviation v(front vehicle, speed Limit)
enable neg. acceleration
scan of surroundings out of road
warning in case of v-deviation
front vehicle: keep speed constant
control speed of Ego-Vehicle
judge speed of front vehicle
front vehicle: enable pos. acceleration
inform about v(front vehicle)
inform about v(Ego-vehicle)
provide speed limit
front vehicle: enable neg. acceleration
enable pos. acceleration

Fig. 16. Example of potential effects of the function *to monitor traffic area*, if the timing variability is changed in comparison to the scenario in Fig. 15. The affected functions are highlighted in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of Anfara et al. (2002) could be a promising approach in assessing the validity of a FRAM model.

Last but not least, the application of FRAM to the overtaking scenario per se is discussed in detail. The functions and their variabilities were identified more or less subjectively. This will have to be based on more objective and empirical research in future to obtain more realistic results. The functional decomposition of the road system by Kuzminski et al. (1995) and the application of SAFE (Fastenmeier & Gstalter, 2007) in particular did not provide sufficient material as a basis to start with FRAM. In future, this should be enhanced using the driving task descriptions in McKnight & Adams (1970) or the reworked and adapted version in Walker & Stanton (2017). Moreover, the graphical representation of the FRAM model in Section 3.2.2 shows that it is in fact possible to visualise the mechanisms of a complex process like the overtaking scenario in road traffic on a functional level. Apart from that, the quantified outputs in Section 3.2.4 (see Figs. 11 and 12) demonstrate their potential to illustrate the critical functions and also to compare these functions for two different performing agents. Besides, the numerical figures provide a quick and comprehensive overview of complex processes within the system. They show the critical functions that then have to be analysed in more detail in the qualitative mapping of the system in a combination of statistical outputs to explore the different mechanisms and emergent processes (see Figs. 14 to 16). The representation of the *GSV* in Fig. 13 has to be treated with caution, as mentioned above: the single weighting and robustness factors of each function in particular will have to be defined in empirical manner in future research.

There are also some general issues regarding the modelling by FRAM in this paper. First of all, the question arises as to whether the complete overtaking manoeuvre can be divided up into three individual models because, for example, functions from model 2 can have a direct influence on other functions from model 3, but this is not shown in this work. An entire model may have to be created in future or the functions with cross-model dependencies may also have to be included in the individual models. The second aspect belongs to the modelling focus. The input used to identify the functions is rather oriented to action sequences of the driving scenario. This has lead that the modelling focus is more on the actions and less on the processes of information processing and decision-making which underlying the actions. The analysis in Sections 3.2.5 shows that these processes are the foundation for safe vehicle guidance and often the most critical aspects dealing with variability and uncertainty. This can be interpreted as a call to shift the modelling focus but at least to consider this in the FRAM model in future research. In general, this would be applicable to both the human driver and automated control system. Regarding the applicability of FRAM, this has the following consequences: these functions are the mechanisms of the inner workings of a human being and automation system and this complicates their access (observability) as Hutchins (1995) emphasising that researchers have easier access to the external representations in a system than to the internal representations resulting from processes within individual actors. Hence, Henriqson et al. (2011) concluded that from a co-agency perspective, human or even artificial cognition can only be comprehend through emergent phenomena of local interactions. Further, Adriaensen et al. (2019) basically discussed that in a joint cognitive system perspective it is better to avoid unnecessary mental constructs, rather focusing on agency even in terms of technical artefacts. Additionally, Adriaensen et al. (2017) suggest being mindful to use FRAM functions that would represent possible opaque cognitive processes. The reason is that these opaque cognitive processes can lead to non-events (nothing observable happens) but there is still a hidden process. For example, in the case of judging the distance to a lead vehicle there can be a possible set of reactions that the driver undertakes, such as braking, going of the throttle, accelerating or steering. Another possible reaction is that the driver does nothing and this could be either because there is no need to react or due to an error of judgement. This "no-reaction" should be modelled by a FRAM function enabling to assess the absence of a signal. In the future, the aim should be to eliminate as many opaque constructs as possible by observable effects, whereby this requires caution with regard to the cause-effect relationships.

However, it would be challenging to define the variability of these functions by more objective and empirical measures. Maybe this requires thought-out interview techniques, eye-tracking as a "window to the mind" or even a neuro-ergonomics approach (Parasuraman, 2011). The eye-mind assumption means that the eye fixates objects whose internal representations are also being processed (Just & Carpenter, 1976). So, the distribution of visual attention can be used as an indicator for cognitive processes. Additionally, specifications in the literature of such processes for the human driver can be used to derive the variability. In contrast, these approaches cannot be applied for the automation and here there seems to be no alternative than SME's. The reason for this is a general data poverty or its public access, although it is generally easier to determine cognitive functions in an automated system than in humans due to the physical architecture in software and hardware terms. Whereas, deep learning and their self-learning algorithms would impair understanding again. In the end, this means a discrepancy between WAD and WAI, since the generation of a model in terms of WAD is almost possible for the human driver but at this moment nearly impossible for the automation. All in all, this shift in modelling focus reveals new methodological issues which exacerbate the problem as mentioned above to identify system functions and their interdependencies as well as their variabilities.

Furthermore, the models or maps created by FRAM are the same for the human driver and the automation, assuming that there is no change in the functions of the system that have to be accomplished by the human driver or the automation. The only difference between the two agents is the variable performance of each function. The reason is that a FRAM model should treat humans and automation systems as equivalent producers of functions to compare the joint performance of both systems as the net result of the functional resonances as depicted by the *GSV*. This may apply to most functions, especially the foreground functions, but there may be other functions, especially background functions, that differ and should be mapped accordingly. According to the authors, the differences depend largely on the abstraction of the functions. The more detailed and specific the description of functions, the more differences arise in the functional models between human drivers and automation. This may have to be considered in future FRAM modelling to understand the real qualitative differences between human drivers and automation.

Moreover, FRAM was applied to a relatively simple overtaking manoeuvre on a rural road in this work, but it is nevertheless complex enough to analyse the interaction between traffic participants and technical systems. However, the question arises as to whether FRAM is applicable for large-scale complex scenarios such as busy city intersections, where vulnerable road users are also brought into the equation. According to Hollnagel (2017), this should be possible because FRAM is basically capable of describing any type of activity or system, be it ever so extraordinary or complex. Even the FRAM method itself can be described by means of FRAM.

Finally, it should be mentioned that the resulting FRAM models do not claim to be complete or highly concretised. Only the most important functions should be presented in order to evaluate FRAM in a purely methodical way and no substantive results regarding the safety contribution of the human driver or automation should be derived. Also, the rating of FRAM functions which was done by the authors should be iterated by further experts. However, the model is a very good starting point and needs to be further specified in the future. Overall, the created FRAM model is intended as a "toy-model", which offers a lot of opportunities for expansion and differentiation.

It can be concluded that the first application of FRAM in the context of the road system demonstrates its suitability to provide information for designing automated driving systems, namely which functions or sequences regarding driving tasks should be automated or not and how these systems should be designed to be safe and effective. FRAM can also be used to reduce the validation effort by providing exclusion criteria, highlighting the critical functions that have to be validated before all other. Especially, FRAM offers a deep understanding of the road system mechanisms which helps to reveal hidden risks of automated driving. Nevertheless, there are some methodological issues that will have to be improved in future research and are summarised in the outlook.

5. Conclusion and outlook

This paper reveals that FRAM is a suitable method to differentiate between the mechanisms of road traffic and to identify the interdependencies between each system element so as to finally identify the contribution of the human driver to road traffic safety within specific situations and to derive requirements for the automation and the potential of automation in these situations with its accompanying factors.

Thus, FRAM should be used as a supporting tool to deliver recommendations for the design of automated driving systems and, in addition, to reduce the validation effort. It should even be mentioned that not only can the mechanisms in road traffic be represented separately for the human driver or the automation, but also the interaction or cooperation between both agents in FRAM. Therefore, the use of FRAM is as well recommended to evaluate ADAS, see also for example the assessment of operational impacts of automation in air traffic by Ferreira and Cañas (2019).

Additionally, the application of FRAM in this paper offers an opportunity to compare this method with the already to road traffic applied STAMP/STPA and AcciMap approaches mentioned in Sections 2.4.1 and 2.4.2. This closes a small research gap regarding the comparison of these methods in terms of safety assessment of automated driving.

Nevertheless, there are some issues that need improvement in future research. The identification of functions and their variability have to be improved in further research, including more objective and empirical measures. This could be based on driving simulator studies, including driving data such as speed, acceleration or distances to other road users, eye-tracking data to observe scanning behaviour and cognitive processes (cf. Arenius, 2017), and interviews for subjective and additional data. Especially, data from sensor technologies can support the traditional qualitative inputs with regard to the following aspects: temporal resolution, gradual differences, time-stamped data and continues recording, coverage and calibration (Arenius, 2017). This is particularly interesting in the context of the very dynamic and complex driving task in road traffic. Otherwise, this will certainly create large amounts of data that require the use of automated data analysis. Additionally, traffic data related to accident black spots or near misses with respect to both the human driver and the automation can be used to provide information for variability identification as well as definition. Moreover, obtained driver performance data such as the compiled data in the Driver Performance Data Book from Henderson et al. (1987) could be used. Overall, it is recommendable to use multiple data in order to integrate these multiple limited perspectives as a mixture of qualitative and quantitative inputs to understand more in-depth the mechanisms and workings of a system. It is relatively easy to apply this approach to generate new data or use existing data related to the human driver. Unfortunately, gathering data related to the automation is complicated for two reasons. Firstly, there are only a few automated vehicles with SAE level 3 or higher on the road for test purposes, so there is little test data. Additionally, much of the data that is generated is kept internally and is not published. Furthermore, when data or information is published, such as the annual Autonomous Vehicle Disengagement Reports from DMV California, this is often not sufficient to draw adequate and meaningful conclusions. The second aspect concerns testing a system directly in a simulation. This requires a complete system on the part of the software and hardware, which is not available. One solution could be to investigate the state-of-the-art of assistance systems and autonomous systems and to collect the performance data from all system components, such as Lidar, cameras or image processing algorithms, in an overview for certain scenario parameters.

Another opportunity for future research, after creating a model with its functions and variabilities for a specific scenario, is to estimate the influence of certain conditional factors (sources of external variability) within the same scenario. This could be achieved by determining an influence factor of the respective conditional factor on each function within the model. For example, this could be used to understand the potential impact of weather-related factors such as fog or sudden events such as wildlife traversal in the whole system. Thus, one can determine how many functions would be affected, and how does the *GSV* behave. Ultimately, the criticality of these conditional factors in the overall system can be considered for different system designs.

Following these methodological improvements, the objective will be to apply FRAM to specific traffic scenarios and to derive results that allow a proactive assessment of the consequences of automation. The resulting recommendations for automated system design and effective measures should enable greater safety in the overall road system.

In conclusion, the safety challenge as a result of automated driving requires tools that take into account high variability and uncertainty. In particular, we will need a safety-II perspective and data of everyday performance of the driving task to understand why things usually go right and sometimes wrong. Additionally, the required tools should enable to study the interactions and mechanisms of a system and practitioners have to consider safety holistically due to the complex and dynamic nature of the road system. FRAM considers these issues and could be the missing piece in the puzzle for a risk assessment of automated driving as well as its system design, and thus according to Ferreira and Cañas (2019):

> “Focus must shift from the streamlining of processes, towards recognising the inevitable need to cope with variability and uncertainty, as they are the means through which complex human endeavours can be achieved. No other element in a system copes better with variability and uncertainty than the human. Technology should, therefore, be addressed as additional resources to cope with increased system capacity, as opposed to a replacement of human resources.”

Acknowledgements

We would like to thank the editing service KERN for proof reading the paper.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Abdulkhaleq, A., Baumeister, M., Böhmert, H., Wagner, S., 2018. Missing no Interaction—Using STPA for Identifying Hazardous Interactions of Automated Driving Systems. Int. J. Saf. 2 (01), 115–124.

Abdulkhaleq, A., Lammering, D., Wagner, S., Röder, J., Balbierer, N., Ramsauer, L., Boehmert, H., 2017. A systematic approach based on STPA for developing a dependable architecture for fully automated driving vehicles. Procedia Eng. 179, 41–51.

Adriaensen, A., Patriarca, R., Smoker, A., Bergström, J., 2019. A socio-technical analysis of functional properties in a joint cognitive system: a case study in an aircraft cockpit. Ergonomics 1–19.

Adriaensen, A., Patriarca, R., Smoker, A., Bergström, J., 2017. Can artefacts be analyzed as an agent by itself–yes or no: what does Hutchins ‘how does a cockpit remember its speeds’tell us. In: 7th REA Symposium, p. 20.

Alvarez, S., 2017. Safety benefit assessment, vehicle trial safety and crash analysis of automated driving: A Systems Theoretic approach (Doctoral dissertation).

Anfara Jr, V.A., Brown, K.M., Mangione, T.L., 2002. Qualitative analysis on stage: Making the research process more public. Educ. Researcher 31 (7), 28–38.

Arenius, M., 2017. Identification of Change Patterns for the Generation of Models of Work-as-Done using Eye-tracking (Vol. 22). Kassel University Press GmbH.

Belmonte, F., Schön, W., Heurley, L., Capel, R., 2011. Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model: An application to railway traffic supervision. Reliab. Eng. Syst. Saf. 96 (2), 237–249.

Bengler, K., Winner, H., Wachenfeld, W., 2017. No Human–No Cry? at-Automatisierungstechnik 65 (7), 471–476.

Bridges, K.E., Corballis, P.M., Hollnagel, E., 2018. “Failure-to-Identify” Hunting Incidents: A Resilience Engineering Approach. Hum. Factors 60 (2), 141–159.

Bubb, H., Bengler, K., Grünen, R.E., Vollrath, M., 2015. Automobilergonomie. Springer-Verlag.

Dallat, C., Salmon, P.M., Goode, N., 2017. Risky systems versus risky people: To what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature. Safety Sci.

Das, S., Sun, X., Wang, F., Leboeuf, C., 2015. Estimating likelihood of future crashes for crash-prone drivers. J. Traffic Transport. Eng. (English edition) 2 (3), 145–157.

De Carvalho, P.V.R., 2011. The use of functional resonance analysis method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience. Reliab. Eng. Syst. Saf. 96 (11), 1482–1498.

Dekker, S., Cilliers, P., Hofmeyr, J.H., 2011. The complexity of failure: Implications of complexity theory for safety investigations. Saf. Sci. 49 (6), 939–945.

Destatis, 2017. Verkehrsunfälle - Fachserie 8 Reihe 7 - 2017 (version from 16.08.2018). Access at 02.01.2019.

Destatis, 2016. Unfallentwicklung auf deutschen Straßen 2015. Wiesbaden. Available at www.destatis.de.

De Winter, J.C.F., Hancock, P.A., 2015. Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them? Procedia Manuf. 3, 5334–5341.

DMV California (Ed.). Access at https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing [12.07.2019].

Eurocontrol, A., 2009. White Paper on Resilience Engineering for ATM. Report of the Project Resilience Engineering for ATM.

Fastenmeier, W., 2015. Fahrerassistenzsysteme (FAS) und Automatisierung im Fahrzeug–wird daraus eine Erfolgsgeschichte. Zeitschrift für Verkehrssicherheit 61 (1), 15–25.

Fastenmeier, W., Gstalter, H., 2007. Driving task analysis as a tool in traffic safety research and practice. Saf. Sci. 45 (9), 952–979.

Ferreira, P.N., Cañas, J.J., 2019. Assessing operational impacts of automation using functional resonance analysis method. Cogn. Technol. Work 1–18.

Frost, B., Mo, J.P., 2014, May. System hazard analysis of a complex socio-technical system: the functional resonance analysis method in hazard identification. In: Proc. of Australian System Safety Conference, Melbourne Australia, pp. 28–30.

Gründl, M., 2005. Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial und die Gestaltung von Fahrerassistenzsystemen (Doctoral dissertation).

Heinrich, H.W., 1941. Industrial Accident Prevention. A Scientific Approach. Industrial Accident Prevention. A Scientific Approach, (Second Edition).

Heinrich, S., Pöppel-Decker, M., Schönebeck, S., Ulitzsch, M., 2010. Unfallgeschehen auf Landstraßen: eine Auswertung der amtlichen Straßenverkehrsunfallstatistik.

Henderson, R.L., Edwards, M.L., United States, 1987. Driver performance data book. U.S. Dept. of Transportation, National Highway Traffic Safety Administration, Washington, D.C..

Henriqson, E., van Winsen, R., Saurin, T.A., Dekker, S.W., 2011. How a cockpit calculates its speeds and why errors while doing this are so hard to detect. Cogn. Technol. Work 13 (4), 217–231.

Hill, R., Hollnagel, E., 2016. Instructions for use of the FRAM model visualiser (FMV). Received from http://functionalresonance.com/onewebmedia/FMV_instructions_0.4.0.pdf [11.07.2019].

Hollnagel, E., 2019. FRAM: Setting the Scene. Received from http://functionalresonance.com/framily-meetings/framily%202019.html [11.07.2019].

Hollnagel, E., 2018a. Safety-I and safety-II: the past and future of safety management. CRC Press.

Hollnagel, E., 2018b. The Functional Resonance Analysis Method. A brief Guide on how to use the FRAM. Received from http://functionalresonance.com/onewebmedia/FRAM%20Handbook%202018%20v4.pdf [11.07.2019].

Hollnagel, E., 2017. FRAM: the functional resonance analysis method: modelling complex socio-technical systems. CRC Press.

Hollnagel, E., 2009. The ETTO principle: Why things that go right sometimes go wrong. Ashgate, Farnham, UK.

Hollnagel, E., 2004. Barriers and accident prevention Ashgate. Hampshire.

Hollnagel, E., 1999. Accident analysis and barrier functions. Institute for Energy Technology, Halden, Norway.

Hollnagel, E., 1998. Cognitive reliability and error analysis method (CREAM). Elsevier.

Hollnagel, E., Hounsgaard, J., Colligan, L., 2014. FRAM – the Functional Resonance Analysis Method – a handbook for the practical use of the method. Received from http://functionalresonance.com/onewebmedia/FRAM_handbook_web-2.pdf [11.07. 2019].

Hollnagel, E., Pruchnicki, S., Woltjer, R., Etcher, S., 2008. Analysis of Comair flight 5191 with the functional resonance accident model. In: 8th International symposium of the Australian aviation psychology association.

Hollnagel, E., Speziali, J., 2008. Study on Developments in Accident Investigation Methods: A Survey of the" State-of-the-art.

Hughes, B.P., Anund, A., Falkmer, T., 2016. A comprehensive conceptual framework for road safety strategies. Accid. Anal. Prev. 90, 13–28.

Hughes, B.P., Newstead, S., Anund, A., Shu, C.C., Falkmer, T., 2015. A review of models relevant to road safety. Accid. Anal. Prev. 74, 250–270.

Huß, C., 1999. Intelligent Speed Adaption. ISM-Workshop 22.9.1999, Bundesanstalt für Straßenwesen. Bergisch-Gladbach.

Hutchins, E., 1995. How a cockpit remembers its speeds. Cogn. Sci. 19 (3), 265–288.

Jackson, S., 2009. Architecting resilient systems: Accident avoidance and survival and recovery from disruptions (Vol. 66). John Wiley & Sons.

Just, M.A., Carpenter, P.A., 1976. Eye fixations and cognitive processes. Cogn. Psychol. 8 (4), 441–480.

Kuzminski, P., Eisele, J.S., Garber, N., Schwing, R., Haimes, Y.Y., Li, D., Chowdhury, M., 1995. Improvement of Highway Safety I: Identification of Causal Factors Through Fault-Tree Modeling 1. Risk Anal. 15 (3), 293–312.

Laaraj, N., Jawab, F., 2018, April. Road accident modeling approaches: literature review. In: 2018 International Colloquium on Logistics and Supply Chain Management (LOGISTIQUA). IEEE, pp. 188–193.

Larsson, P., Dekker, S.W., Tingvall, C., 2010. The need for a systems theory approach to road safety. Saf. Sci. 48 (9), 1167–1174.

Lemmer, K. (Ed.). (2016). Neue autoMobilität: Automatisierter Straßenverkehr der Zukunft. Herbert Utz Verlag.

Leveson, N., Thomas, J.O.H.N., 2018. STPA handbook. NANCY LEVESON AND JOHN THOMAS, 3.

Leveson, N.G., 2011. Applying systems thinking to analyze and learn from events. Saf. Sci. 49 (1), 55–64.

Leveson, N., 2004. A new accident model for engineering safer systems. Saf. Sci. 42 (4), 237–270.

Leveson, N., 1995. Safeware: System Safety and Computers. Addison-Wesley, Reading, MA.

Lundblad, K., Speziali, J., Woltjer, R., Lundberg, J., 2008. FRAM as a risk assessment method for nuclear fuel transportation. In: Proceedings of the 4th International Conference Working on Safety (Vol. 1, S. 223-1).

Maier, F., 2013. Wirkpotentiale moderner Assistenzsysteme und Aspekte ihrer Relevanz für die Fahrausbildung (Doctoral dissertation, Dissertation, Institute of Ergonomics, Technische Universität München).

Maurer, M., Gerdes, J.C., Lenz, B., Winner, H., 2016. Autonomous driving. Berlin, Germany: Springer Berlin Heidelberg, 10, 978–973.

McIlroy, R.C., Plant, K.L., Stanton, N.A., 2018, August. Revealing the Complexity of Road Transport with Accimaps. In: Congress of the International Ergonomics Association. Springer, Cham, pp. 80–89.

McKnight, A.J., Adams, B.B., 1970. Driver education task analysis. Volume 1: Task descriptions.

Parasuraman, R., 2011. Neuroergonomics: Brain, cognition, and performance at work. Curr. Directions Psychol. Sci. 20 (3), 181–186.

Patriarca, R., Bergström, J., 2017. Modelling complexity in everyday operations: functional resonance in maritime mooring at quay. Cogn. Technol. Work 19 (4), 711–729.

Patriarca, R., Bergström, J., Di Gravio, G., 2017a. Defining the functional resonance analysis space: Combining Abstraction Hierarchy and FRAM. Reliab. Eng. Syst. Saf. 165, 34–46.

Patriarca, R., Del Pinto, G., Di Gravio, G., Constantino, F., 2018. FRAM for systemic accident analysis: a matrix representation of functional resonance. Int. J. Reliab. Qual. Saf. Eng. 25 (01), 1850001.

Patriarca, R., Di Gravio, G., Costantino, F., 2017b. A Monte Carlo evolution of the Functional Resonance Analysis Method (FRAM) to assess performance variability in complex systems. Saf. Sci. 91, 49–60.

Patriarca, R., Di Gravio, G., Costantino, F., 2017c. myFRAM: An open tool support for the functional resonance analysis method. In: 2017 2nd International Conference on System Reliability and Safety (ICSRS). IEEE, pp. 439–443.

Perrow, C., 1984. 1984: Normal accidents: living with high-risk technologies. Basic Books, New York.

Qureshi, Z.H., 2007, December. A review of accident modelling approaches for complex socio-technical systems. In: Proceedings of the twelfth Australian workshop on Safety critical systems and software and safety-related programmable systems-Volume 86. Australian Computer Society, Inc., pp. 47-59.

Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. Saf. Sci. 27 (2–3), 183–213.

Reason, J., 1990. Human error. Cambridge University Press.

Reichart, G., 2000. Menschliche Zuverlässigkeit beim Führen von Kraftfahrzeugen–Möglichkeiten der Analyse und Bewertung (Doctoral dissertation, Dissertation am Lehrstuhl für Ergonomie der TU München).

SAE J3016, 2014. Taxonomy and Definitions for Terms Related to On-Road Motor Vehicle Automated Driving Systems.

Salmon, P.M., McClure, R., Stanton, N.A., 2012. Road transport in drift? Applying contemporary systems thinking to road safety. Saf. Sci. 50 (9), 1829–1838.

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. Accid. Anal. Prev. 96, 140–151.

Schnieder, E., Schnieder, E., 2013. Verkehrssicherheit. Springer Verlag.

Scott-Parker, B., Goode, N., Salmon, P., 2015. The driver, the road, the rules… and the rest? A systems-based approach to young driver road safety. Accid. Anal. Prev. 74, 297–305.

Shirali, G.A., Ebrahipour, V., Mohammd Salahi, L., 2013. Proactive risk assessment to identify emergent risks using functional resonance analysis method (fram): a case study in an oil process unit. Iran Occup. Health 10 (6).

Shladover, S.E., Nowakowski, C., 2017. Regulatory challenges for road vehicle automation: Lessons from the california experience. Transport. Res. Part A: Policy Practice.

Slater, D., 2016. QUANTITATIVE VARIABILITY IN FRAM. Received from https://www.researchgate.net/publication/305886439_QUANTITATIVE_VARIABILITY_IN_FRAM [26.10.2019]. DOI: 10.13140/RG.2.1.2985.7529.

Smoczyński, P., Kadziński, A., Gill, A., Kobaszyńska-Twardowska, A., 2018. Applicability of the functional resonance analysis method in urban transport. In: MATEC Web of Conferences (Vol. 231, p. 05006). EDP Sciences.

Stanton, N.A., Salmon, P.M., Walker, G.H., Stanton, M., 2019. Models and methods for collision analysis: A comparison study based on the Uber collision with a pedestrian. Saf. Sci. 120, 117–128.

Steen, R., Aven, T., 2011. A risk perspective suitable for resilience engineering. Saf. Sci. 49 (2), 292–297.

Svedung, I., Rasmussen, J., 2002. Graphic representation of accident scenarios: Mapping system structure and the causation of accidents. Saf. Sci.

Tian, J., Wu, J., Yang, Q., Zhao, T., 2016. FRAMA: a safety assessment approach based on Functional Resonance Analysis Method. Saf. Sci. 85, 41–52.

Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., Maurer, M., 2015, September. Defining and substantiating the terms scene, situation, and scenario for automated driving. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. IEEE, pp. 982–988.

Underwood, P., 2013. Examining the systemic accident analysis research-practice gap (Doctoral dissertation, © Peter Underwood).

Vicente, K.J., 1999. Cognitive work analysis: Toward safe, productive, and healthy computer-based work. CRC Press.

Visser, E., Pijl, Y.J., Stolk, R.P., Neeleman, J., Rosmalen, J.G., 2007. Accident proneness, does it exist? A review and meta-analysis. Accid. Anal. Prev. 39 (3), 556–564.

Wachenfeld, W., Winner, H., 2016. The release of autonomous vehicles. In: Autonomous driving. Springer, Berlin, Heidelberg, pp. 425–449.

Walker, G.H., Stanton, N.A., 2017. Human factors in automotive engineering and technology. CRC Press.

Wienen, H.C.A., Bukhsh, F.A., Vriezekolk, E., Wieringa, R.J., 2017, June. Accident analysis methods and models—a systematic literature review. In: Centre for Telematics and Information Technology (CTIT).

Winkle, T., 2016a. Safety benefits of automated vehicles: Extended findings from accident research for development, validation and testing. In: Autonomous driving. Springer, Berlin, Heidelberg, pp. 335–364.

Winkle, T., 2016b. Development and approval of automated vehicles: considerations of technical, legal, and economic risks. In: Autonomous Driving. Springer, Berlin, Heidelberg, pp. 589–618.

Winner, H. (2015). Quo vadis, FAS?. In: Handbuch Fahrerassistenzsysteme. Springer Vieweg, Wiesbaden, pp. 1167–1186.

Woltjer, R., Hollnagel, E., 2008. Functional modeling for risk assessment of automation in a changing air traffic management environment. In: Proceedings of the 4th International Conference Working on Safety (Vol. 30).

Yang, Q., Tian, J., Zhao, T., 2017. Safety is an emergent property: Illustrating functional resonance in air traffic management with formal verification. Saf. Sci. 93, 162–177.

Young, K.L., Salmon, P.M., 2015. Sharing the responsibility for driver distraction across road transport systems: a systems approach to the management of distracted driving. Accid. Anal. Prev. 74, 350–359.

C Article 2: “Safety Enhancement by Automated Driving: What are the Relevant Scenarios?”

Grabbe, N., Höcher, M., Thanos, A., & Bengler, K. (2020). Safety Enhancement by Automated Driving: What are the Relevant Scenarios? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 64, No. 1, pp. 1686-1690). Sage CA: Los Angeles, CA: SAGE Publications. https://doi.org/10.1177/1071181320641409

Safety Enhancement by Automated Driving: What are the Relevant Scenarios?

Niklas Grabbe, Michael Höcher, Alexander Thanos and Klaus Bengler
Technical University of Munich – Chair of Ergonomics

Automated driving offers great possibilities in traffic safety advancement. However, evidence of safety cannot be provided by current validation methods. One promising solution to overcome the approval trap (Winner, 2015) could be the scenario-based approach. Unfortunately, this approach still results in a huge number of test cases. One possible way out is to show the current, incorrect path in the argumentation and strategy of vehicle automation, and focus on the systemic mechanisms of road traffic safety. This paper therefore argues the case for defining relevant scenarios and analysing them systemically in order to ultimately reduce the test cases. The relevant scenarios are based on the strengths and weaknesses, in terms of the driving task, for both the human driver and automation. Finally, scenarios as criteria for exclusion are being proposed in order to systemically assess the contribution of the human driver and automation to road safety.

INTRODUCTION

The current test concept for safety approval of motor vehicles is based on a track distance and statistical approach. This means that the vehicle must cover a required test distance under representative conditions in real traffic without accidents. If we keep this test concept for vehicles with SAE-level 3 or higher (automated driving, AD) according to SAE J3016 (2018), then according to Wachenfeld & Winner (2016) about 6.6 billion test kilometres would have to be covered. Thus, the current test methods are not suitable as proof of safety for economical and practical reasons. Therefore, research is being done on alternative approval methods.

One promising solution could be the scenario-based approach. The assumption is that the long test-driving distances needed for statistical validation could be significantly reduced by identifying crucial scenarios that can be reproduced in simulation or in test fields. Unfortunately, this approach leads to a parameter space explosion due to the level of scenario abstraction (Amersbach et al., 2019), which still results in a huge number of test cases.

The scenario generation represents a sensitive step for the safety reasoning of the scenario-based approach. So, what can we do to facilitate this approach and reduce the approval effort? If we take a closer look at the frequent argumentation and strategy of automation, then we can reveal a fallacy which still contains a possible way out as a solution.

Current fallacy and a new perspective

The argument for increased automation in the driving task is often accompanied by the argument that humans, in their role as drivers and the main cause of accidents, could be removed from the system. Consequently, the number of accidents would fall sharply. This argumentation does not take into account that accidents are rare, Poisson-distributed and multi-causal events. Thus, the human driver is not the only cause; in fact in many situations he or she is the essential accident avoidance and compensation element in the system, which also has to be considered. Additionally, the current approach to automating vehicle guidance is a selective strategy. For certain traffic situations and the associated tasks, a classic human driver-vehicle system is still selected due to the lack of automation, for other situations in which this is technically feasible, a human-machine system is based on automation. For this assignment, there should be a correlation between the selection of the respective system characteristic of the automated system and the success criterion "accident-free driving". For example, accident black spots for urban intersections, especially left turns or rural roads, can be identified (Maier, 2013; Gründl, 2005). In addition, certain groups are at increased risk due to misconduct (Das et al., 2015). However, current approaches to vehicle automation address relatively safe traffic situations, such as highway scenarios, and are not directed at risk groups of human drivers. Two aspects are therefore erroneously assumed: the potential for automation is the same in every scenario, and all drivers profit equally from automation and deal with potential side effects of automation in the same way. In fact, above all, the mechanisms of accident development and accident avoidance should be understood in order to assess the driver's contribution in the corresponding situations and thus to derive requirements for automation and the potential of automation with its accompanying factors. (Bengler et al., 2017)

Grabbe et al. (2020) support this statement by recommending a systemic approach to the design and safety assessment of AD (cf. Larsson et al., 2010; Hughes et al., 2015), and in particular the authors revealed that the functional resonance analysis method by Hollnagel (2012) is a suitable method for differentiating between the mechanisms of road traffic. Ultimately, the test cases could be reduced based on key insights from this systemic analysis in crucial scenarios. Before these mechanisms can be analysed, the essential scenarios to be examined must be defined. To select these scenarios, according to Bengler et al. (2017) we establish that scenarios should not only refer to rarely critical events or even more rarely occurring accidents, but that uncritical and accident-free situations must be considered. This is also in accordance to a safety-II perspective (Hollnagel, 2018) that seems urgently required to overcome the approval trap (Grabbe et al., 2020). In addition, scenarios must be considered that seem likely to be critical against the background of the technical characteristics of automation. These are, for example, situations with strong cooperation behaviour, situations in which the errors of other road users have to be compensated for and situations in which events occur outside the sensor range and actions must be anticipated (Bengler et al., 2017).

Objective

The purpose of this paper is to derive basic scenarios as criteria for exclusion, based on the strengths and weaknesses, in terms of the driving task, for both the human driver and automation in order to systemically assess the contribution of the human driver and the automation to road safety.

METHOD

The basic aim is to test scenarios that enable comparison between a purely manual human driver and a fully automated system in a mixed traffic situation including different levels of automation and traffic participants. This means that scenarios in which an interaction between driver and automation takes place, for example in a takeover situation or during assisted driving, are not considered. Rather, scenarios are selected and divided into two different types following the aforementioned suggestions by Bengler et al. (2017).

Type I describes scenarios that offer great potential for a significant safety improvement through automation because humans have proven highly likely to contribute to accident occurrence. In order to identify these scenarios, accident statistics from Germany from 2018 (Destatis, 2019) were analysed to derive accident black spots and risk groups of human drivers. Accidents with material damage are not considered as part of this, but only injuries or fatalities, since these are the most serious in terms of safety. Furthermore, it is assumed that the accident statistics for Germany are also comparable to many other countries and thus the results can be transferred.

Type II comprises scenarios that represent either unique strengths of the human driver in uncritical and accident-free situations, or supposed challenges for automation. To identify these scenarios, a succinct synthesis of literature was performed, and due to the lack of data on accidents or disengagements involving automated vehicles from the department of motor vehicles in California (DMV California), as well as their superficial reporting, semi-structured interviews were held with six experts from the field of driver assistance and vehicle automation in the German automotive industry. The first part of the interviews covered performance limits of the current systems in vehicle automation with regard to information processing. The second part of the interview dealt with the derivation of scenarios regarding the strengths and weaknesses of vehicle automation on highways, in the city and on rural roads. In addition, the experts were asked which scenarios offer the greatest potential for a significant increase in traffic safety through automation and which scenarios represent the supposedly most relevant test scenarios for comparison between vehicle automation and the human driver. The interviews were evaluated using qualitative content analysis through systematic category formation and quantification.

SCENARIOS OF TYPE I

Considering the accidents according to their location in relation to frequency and severity, accidents in the city and on rural roads are by far the most critical. In addition, young drivers up to the age of 25 and older drivers above the age of 65 make up a significantly greater proportion than middle-aged drivers of those causing accidents than those merely involved. (Destatis, 2019) Thus, scenarios in the city, on rural roads and also for risk groups should be highlighted.

Accident black spots on city roads

In the city, collisions with another vehicle that turns or crosses and collisions between vehicles and pedestrians are by far the most common types of accidents. (Destatis, 2019) Since the former type of accident inevitably occurs at a node and a large part of the latter type of accident is also attributable to a nodal area, scenarios in urban areas at intersections or junctions represent the greatest risk factor.

At intersections, the largest proportion of accidents with oncoming vehicles occur when the vehicle is turning left. Other focal points at intersections are conflicts with crossing vehicles from the right and left. In contrast, accidents at junctions are dominated by collisions between vehicles and cyclists or pedestrians crossing on the right. Also, vehicles crossing from the left represent a further focus. A look at the right-of-way rule shows that almost two thirds of all intersection accidents occur at traffic light systems, and instead in the case of accidents at junctions, a significant number of accidents occur at sign-regulated junctions. The proportion of accident sites with "right before left" regulation is approximately twice as large in intersection accidents than in accidents at junctions. (Gerstenberger, 2015)

Finally, type I scenarios should include urban scenarios at a traffic light-controlled intersection with a left turn and oncoming traffic, at an intersection with right before left and at a sign-regulated junction with vulnerable road users crossing from the right or a vehicle crossing from the left.

Accident black spots on rural roads

On rural roads, collisions with oncoming vehicles and leaving the carriageway pose the greatest danger (Destatis, 2019). By far the largest proportion of collisions with oncoming vehicles is caused by overtaking manoeuvres (Richter & Ruhl, 2014). Overtaking manoeuvres also account for an increased proportion of accidents when vehicles leave the carriageway, since dangerous evasive manoeuvres in overtaking situations often cause vehicles to leave the carriageway. However, the main reasons for accidents that result in vehicles leaving the carriageway are excessive speed and neglecting the safety distance. Nevertheless, with regard to the former reasons, overtaking scenarios on rural roads should be emphasised.

The majority of overtaking drivers collide with oncoming traffic; the main reasons for this are overtaking despite oncoming traffic and unclear traffic conditions. Most overtaking accidents occur on straight roads, followed by overtaking accidents on winding road areas, especially in right-hand bends due to poor visibility of this bend in right-hand traffic. Additionally, overtaking accidents increase with a decreasing curve radius. Furthermore, many overtaking accidents occur in the immediate vicinity of hilltops. Lastly, weather influences are of minor importance in overtaking accidents, since most overtaking accidents occur in daylight and on dry roads, all in all under good external conditions. (Richter & Ruhl, 2014)

In summary, scenarios of type I should incorporate rural scenarios with overtaking manoeuvres on a straight road, at a hilltop and in a tight bend to the right, all with oncoming traffic and in good external conditions.

Risk groups

In rural overtaking scenarios, young drivers should be considered due to excessive speed (Destatis, 2019) and driving accident black spots due to their vehicles leaving the carriageway (Gründl, 2005) and overtaking (Richter & Ruhl, 2004). Older drivers should be the focus in urban scenarios at intersections because they are more likely to make right-of-way mistakes (Destatis, 2019) and they have showed greater weaknesses when turning and crossing (Gründl, 2005).

SCENARIOS OF TYPE II

All experts noted that motorways are by far the easiest scenarios to implement, relatively speaking, because the environment is very clear and reproducible due to the highly regulated infrastructure with few different types of road user. On the other hand, city roads and rural roads are currently the greatest flaw in vehicle autonomy. In addition, the unique strengths of human drivers are more likely to be put to use in urban or rural scenarios. Therefore, type II scenarios focus on these two locations. Also, according to experts, the most relevant test scenarios for approval of autonomous vehicles should deal with the following areas: interaction with other road users, complex scenarios with little model structural information involving strong use of knowledge-based behaviour as well as unexpected special situations, different weather and road conditions and overtaking manoeuvres and particularities on rural roads. Thus, the scenarios are classified into five different categories.

Communication and interaction

Färber (2016) describes human beings as multi-sensory, adaptive systems that can understand and interpret various, weak and ambiguous signals. In contrast, machines are restricted to strict rules and cannot understand the informal rules of humans. In road traffic, people communicate not only via prescribed signals such as indicators, brake lights and horns, but also through informal communication channels. According to Merten (1977), various options are available for communication: schema formation, anticipatory behaviour, non-verbal communication, facial expressions, eye contact, gestures and body movements. The automation must also be able to perceive all of these informal signs and interpret them in the environmental context in order to predict the behaviour of others. Based on this, the automation can adopt an adapted behaviour, which must also be understood by other road users. Similarly, the experts expect the automation to exhibit shortcomings on city roads when interacting with vulnerable road users and other motor vehicles due to high complexity, diverse road users and masking objects.

Thus, urban scenarios of type II should include an interaction with a pedestrian or cyclist, an interaction with a motorised vehicle at a two-sided bottleneck, a simultaneous lane change of two vehicles and a zip-merge.

Complexity and anticipation

The human driver is very flexible and strong in heuristic thinking. This means that complex scenarios that cannot be fully described, and those in which not all information is available, are easier to master for humans than for automation. Situations in which rules are broken, errors have to be compensated for and events have to be anticipated are also a strength of the human driver (cf. MABA-MABA list by Fitts, 1951; Winter and Hancock, 2015). Here, the experts see a real bottleneck in current automated systems: generally, driving per se (control of the vehicle) is not the problem, rather it is troublesome when decisions are being made or when information is incomplete, and when something unexpected happens. Thus, the biggest problem currently lies in the level of cognition. This means that the autonomous vehicle has problems interpreting and recognising relationships between individual objects and their meaning. A limited scope for action is particularly difficult due to the highly rule-based design of automation and less through self-learning neural networks. It is difficult to use incomplete knowledge associatively and through heuristic thinking for good interpretation, anticipation and interaction.

Further, the anticipation performance of human drivers in road traffic should be analysed, especially against the background of a limited range of sensors. Above all, for good anticipation performance it is necessary to observe distant characteristics, consider the history of the driving scene and make assumptions about the intentions of other drivers. The anticipation is negatively affected by masking, less salient features and features that are in an area in which no particularly relevant stimuli are expected. (Sommer, 2013)

In summary, type II scenarios should include a complex urban intersection, an anticipation scenario on a city road (e.g. a vehicle in front of the leading vehicle is parking on the right-hand side of the road) and a scenario with failure compensation (e.g. a left bend on a rural road in which an overtaking vehicle has to be avoided).

Special situations

The experts currently see flaws in the recognition of special objects by sensors. Moreover, special situations such as construction sites, accident sites or situations in which rules are not followed by others also represent a shortcoming for vehicle automation, as the autonomous system acts in a very rule-based manner, and therefore has problems with decision logic in situations in which people rely on their intuition, experience and general knowledge.

Thus, urban scenarios of type II should include a construction site, an accident site especially with an emergency lane, and a deliberate violation of the traffic rules by crossing a solid lane onto the oncoming lane due to a broken-down vehicle in its own lane and then interacting with oncoming traffic.

Environmental conditions

Dixit et al. (2016) showed that road infrastructure problems are a common reason for disengagements of AD. Also, the experts see flaws in poor weather conditions. Therefore, rural scenarios of type II with respect to different road/weather conditions and road damage have to be considered. These could encompass a slippery, winding road due to snow and frost and a rural bend with unpaved roadside, interrupted lane markings, potholes and dangerous small parts on the road.

Rural road and overtaking

The experts see a huge problem for autonomous vehicles on rural road especially with regard to the high number of sight obscurations, tight bends, narrow roads, unseparated directional lanes, missing lane and roadside markings and difficult weather conditions. Also, the overtaking manoeuvre in combination of some of these factors should be considered. Thus, type II scenarios could include a winding rural road at night with much masking, without lane markings and a narrow lane, and an overtaking on a straight rural road with oncoming traffic and a turning truck from a junction.

CLASSIFICATION AND ANALYSIS OF RELEVANT SCENARIOS

The relevant scenarios are classified in Table 1 for a better overview with an assignment to the scenario types, categories and important key features. The scenarios currently represent only abstract and basic situations and, regarding the subsequent systemic analysis, the following parameters especially concerning urban scenarios, according to Nambuusi et al. (2008), should be varied: traffic volume, pedestrian/cyclist traffic, lighting conditions, type of right-of-way rule, junction or road geometry, environmental conditions, and vehicle types. Accordingly, conceivable parameters to be varied on rural roads are the following: speed of the vehicle ahead, vehicle speed on the opposite lane, traffic volume and surroundings, speed limit, routing, bends and gradients, weather conditions and visibility, time pressure. Also, the demonstrated risk groups should be analysed in the corresponding scenarios. For a future consistent description of the derived scenarios, it is recommended to use the definitions and terms of Ulbrich et al. (2015).

When analysing the scenarios, it is less important to pay attention to critical events such as errors or accidents than it is to focus on the variability in the performance and the qualitative execution of the individual driving tasks in uncritical and normal driving. This should be assessed in terms of a safety-II perspective and a systemic approach (cf. Grabbe et al., 2020), which also means that significantly fewer test kilometres have to be covered since data can be gathered immediately. There is no need to wait for the accident to occur or something bad to happen, because you can measure anything at any time. Rather, we have to understand what actually happens in situations where nothing out of the ordinary seems to take place and to compare this between the human driver and the automated vehicle. Thus, it is sufficient to develop a description of the daily activity and its expected variability, which means one generic case instead of many specific ones. Therefore, it makes more sense to analyse small but frequent events (everyday performance) instead of large but rare events (accidents), because the former are easier to understand and manage, and also have more impact on the safety of the overall system (cf. Hollnagel, 2018). Clearly this will result in a large amount of data that has to be analysed automatically. Fortunately, this problem should be solvable compared to the extrapolated lots of test kilometres by Wachenfeld & Winner (2016), which is based on track distance that currently focuses only on accidents.

DISCUSSION

This paper presents a new perspective on the scenario-based approach to overcoming the current approval trap

Table 1.Classification of the relevant scenarios

	Scenarios of type I			**Scenarios of type II**				
	A	**B**	**C**	**A**	**B**	**C**	**D**	**E**
Category	Accident black spots in the city	Accident black spots on rural roads	Risk groups	Communication and interaction	Complexity and anticipation	Special situations	Environmental conditions	Rural road and overtaking

Key features	Turning and crossing at urban intersection and junctions	Overtaking on rural straight roads, at hilltops or in tight bends	Overtaking on rural roads;
			turning and crossing at urban intersections and junctions
				Interacting with vulnerable road users or with a motorised vehicle;
				simultaneous lane change of two vehicles;
				zip-merge
					Complex urban nodal areas;
					anticipation in urban and rural areas;
					failure compensation in urban or rural areas
						Urban construction sites;
						urban accident sites;
						deliberate violation of traffic rules on a city road
							Winding rural road in poor weather conditions;
							rural bend with poor road conditions and damage
								Winding rural road at night with much masking, without lane markings and a narrow lane;
								complex overtaking manoeuvre

(Winner, 2015) of automated vehicles. Current approaches and projects for the safety assessment of AD are strongly technology driven, pursue a clear safety-I perspective in the safety argumentation and focus too much on the vehicle itself. Instead, the new perspective in this paper strongly argues for a safety-II perspective and a systemic approach. In particular, relevant scenarios as criteria for exclusion are derived in order to systemically assess (Grabbe et al., 2020) the contribution of human drivers and automation to road safety. Finally, based on the key insights from the subsequent systemic analysis in relevant scenarios, the validation work could be reduced to a reasonable effort by minimising the possible parameter space.

Despite the fact that critical accident scenarios and risk groups among human drivers can be shown very effectively, a methodological limitation is the derivation of human driver's strengths. Past and current data analysis almost always focuses on errors and accidents, but does not consider accident avoidance or, in particular, what went right. In addition, it was difficult to derive the required scenarios with respect to automation, since little data on failures and accidents or the driving behaviour of automation is public. In the case of the data collected by the DMV California, the reports are inadequate and superficial, so no meaningful knowledge can be gained from them. Therefore, in the future, it will be necessary, on one hand, to consider broad-based data collection for human drivers in uncritical and accident-free situations (c.f. Bengler et al., 2017) and, on the other, to make the reporting of disengagements and accidents by automated vehicles significantly more in-depth.

Nevertheless, it is not possible to prove that all relevant scenarios have been captured completely, and the derived scenarios do not claim to be complete. For example, local and cultural characteristics may not all be considered. But with the presented relevant scenarios as decisive factors, we have a strong foundation to systemically analyse these scenarios in order to build an understanding of a system actual mechanisms that are needed to support the design of safe automated vehicles proactively and to reduce the validation work.

In future work, we propose analysing the derived scenarios using the novel approach by Grabbe et al. (2020) in order to reduce the approval effort by a scenario-based approach based on the pre-selection of relevant scenarios and their systemic analysis. This would give us increased knowledge of the systemic interrelationships, which offers great potential to ensure safe automation as well as to reduce the validation effort. The main goal must be to improve the safety of the overall system through the efficient interaction of human drivers, machines and road users among themselves. The question, therefore, is not how can we make vehicle automation safe and prove that it is safe, but instead how can we use automation in order to design a safe traffic system.

REFERENCES

Amersbach, C., & Winner, H. (2019). Functional decomposition - A contribution to over-come the parameter space explosion during validation of highly automated driving. Traffic injury prevention, 20(sup1), S52-S57.

Bengler, K., Winner, H., & Wachenfeld, W. (2017). No Human–No Cry?. at-Automatisierungstechnik, 65(7), 471-476.

Das, S., Sun, X., Wang, F., & Leboeuf, C. (2015). Estimating likelihood of future crashes for crash-prone drivers. Journal of traffic and transportation engineering (English edition), 2(3), 145-157.

Dixit, V. V., Chand, S., & Nair, D. J. (2016). Autonomous vehicles: disengagements, accidents and reaction times. PLoS one, 11(12), e0168054. https://doi.org/10.1371/jour-nal.pone.0168054.

DMV California. Retrieved from https://www.dmv.ca.gov/portal/dmv/detail/vr/autonomous/testing [14.01.2019].

Destatis (2019). Verkehrsunfälle – Fachserie 8 Reihe 7 -2018. Retrieved from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/verkehrsunfaelle-jahr-2080700187004.html.

De Winter, J. C. F., & Hancock, P. A. (2015). Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them?. Procedia Manufacturing, 3, 5334-5341.

Fitts, P. M. (1951). Human engineering for an effective air-navigation and traffic-control system.

Färber, B. (2016). Communication and communication problems between autonomous vehicles and human drivers. In Autonomous driving (pp. 125-144). Springer, Berlin, Heidelberg.

Gerstenberger, M. (2015). Unfallgeschehen an Knotenpunkten (Doctoral dissertation, Technische Universität München).

Grabbe, N., Kellnberger, A., Aydin, B. and Bengler, K. (2020). Safety of Automated Driving: The Need for a Systems Approach and Application of the Functional Resonance Analysis Method. *Safety Science, 126*. DOI: 10.1016/j.ssci.2020.104665.

Gründl, M. (2005). Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial und die Gestaltung von Fahrerassistenzsystemen (Doctoral dissertation).

Hollnagel, E. (2018). Safety-I and safety-II: the past and future of safety management. CRC press.

Hollnagel, E. (2012). FRAM: the functional resonance analysis method: modelling complex socio-technical systems. CRC Press

Hughes, B. P., Newstead, S., Anund, A., Shu, C. C., & Falkmer, T. (2015). A review of models relevant to road safety. Accident Analysis & Prevention, 74, 250-270.

Larsson, P., Dekker, S. W., & Tingvall, C. (2010). The need for a systems theory approach to road safety. Safety science, 48(9), 1167-1174.

Maier, F. (2013). Wirkpotentiale moderner Assistenzsysteme und Aspekte ihrer Relevanz für die Fahrausbildung (Doctoral dissertation, Dissertation, Institute of Ergonomics, Technische Universität München).

Merten, K. (1977). Kommunikationsprozesse im Straßenverkehr. Symposion, 77.

Nambuusi, B., Brijs, T., & Hermans, E. (2008). A review of accident prediction models for road intersections.

Richter, T., & Ruhl, S. (2014). Untersuchung von Maßnahmen zur Prävention von Überholunfällen auf einbahnigen Landstraßen. Unfallforschung der Versicherer, GDV.

SAE International (June 2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. (J3016).

Sommer, K. C. (2013). Vorausschauendes Fahren - Erfassung, Beschreibung und Bewertung von Antizipationsleistungen im Straßenverkehr (Doctoral dissertation).

Ulbrich, S., Menzel, T., Reschka, A., Schuldt, F., & Maurer, M. (2015, September). Defining and substantiating the terms scene, situation, and scenario for automated driving. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems (pp. 982-988). IEEE.

Wachenfeld, W., & Winner, H. (2016). The release of autonomous vehicles. In Autonomous driving (pp. 425-449). Springer, Berlin, Heidelberg.

Winner, H. (2015). Quo vadis, FAS?. In Handbuch Fahrerassistenzsysteme (pp. 1167-1186). Springer Vieweg, Wiesbaden.

D Article 3: “Functional Resonance Analysis in an Overtaking Situation in Road Traffic: Comparing the Performance Variability Mechanisms between Human and Automation”

Grabbe, N., Gales, A., Höcher, M., & Bengler, K. (2022). Functional Resonance Analysis in an Overtaking Situation in Road Traffic: Comparing the Performance Variability Mechanisms between Human and Automation. *Safety, 8*(1), 3. https://doi.org/10.3390/safety8010003

Article

Functional Resonance Analysis in an Overtaking Situation in Road Traffic: Comparing the Performance Variability Mechanisms between Human and Automation

Niklas Grabbe *, Alain Gales, Michael Höcher and Klaus Bengler

Chair of Ergonomics, Technical University of Munich, 85748 Garching, Germany; alain.gales@tum.de (A.G.); michi-hoecher@web.de (M.H.); bengler@tum.de (K.B.)
* Correspondence: n.grabbe@tum.de; Tel.: +49-89-289-15375

Abstract: Automated driving promises great possibilities in traffic safety advancement, frequently assuming that human error is the main cause of accidents, and promising a significant decrease in road accidents through automation. However, this assumption is too simplistic and does not consider potential side effects and adaptations in the socio-technical system that traffic represents. Thus, a differentiated analysis, including the understanding of road system mechanisms regarding accident development and accident avoidance, is required to avoid adverse automation surprises, which is currently lacking. This paper, therefore, argues in favour of Resilience Engineering using the functional resonance analysis method (FRAM) to reveal these mechanisms in an overtaking scenario on a rural road to compare the contributions between the human driver and potential automation, in order to derive system design recommendations. Finally, this serves to demonstrate how FRAM can be used for a systemic function allocation for the driving task between humans and automation. Thus, an in-depth FRAM model was developed for both agents based on document knowledge elicitation and observations and interviews in a driving simulator, which was validated by a focus group with peers. Further, the performance variabilities were identified by structured interviews with human drivers as well as automation experts and observations in the driving simulator. Then, the aggregation and propagation of variability were analysed focusing on the interaction and complexity in the system by a semi-quantitative approach combined with a Space-Time/Agency framework. Finally, design recommendations for managing performance variability were proposed in order to enhance system safety. The outcomes show that the current automation strategy should focus on adaptive automation based on a human-automation collaboration, rather than full automation. In conclusion, the FRAM analysis supports decision-makers in enhancing safety enriched by the identification of non-linear and complex risks.

Keywords: automated driving; human driving; risk assessment; resilience engineering; systems thinking; overtaking manoeuvre

check for updates

Citation: Grabbe, N.; Gales, A.; Höcher, M.; Bengler, K. Functional Resonance Analysis in an Overtaking Situation in Road Traffic: Comparing the Performance Variability Mechanisms between Human and Automation. *Safety* **2022**, *8*, 3. https://doi.org/10.3390/safety8010003

Academic Editor: Raphael Grzebieta

Received: 26 August 2021
Accepted: 17 December 2021
Published: 27 December 2021

1. Introduction

In the past, traffic safety was improved by three major safety strategies including engineering, enforcement, education [1], and their intertwinings. Nevertheless, according to the World Health Organisation [2], over 1.2 million people die each year on the world's roads, and between 20 and 50 million suffer non-fatal injuries. These are still high numbers that need to be improved. A promising countermeasure seems to be a technology advancement by automated driving (AD, Level 3 and higher, according to SAE J3016 [3]), which offers great possibilities in traffic safety enhancement. A frequent argumentation for this assumption is that the human in his role as a driver is the main cause of accidents, claiming that human error causes approximately 90% of road crashes, e.g., [4–7]. Consequently, it is frequently recommended that the human driver be removed from the system and road accidents will probably decrease by 90%. The common idea behind this is that technology can be introduced into a system by simply substituting machines for people so that the system as a whole improves and there are no negative side effects. Unfortunately, this is a persistent oversimplification fallacy, also called substitution myth [8,9].

This is in accordance with Rasmussen [10], who claimed that for a general understanding of system behaviour, we do not have to focus on human errors, rather on mechanisms shaping the behaviour in the system and its context. This is also in line with Woods and Dekker [8], who stated that it is not sufficient to build a new mature system or technology and then to test and assess its performance at the tail of the design process. Rather a proactive approach at the beginning has to steer the design into a direction that considers the usefulness of a potential new system given the possibilities new technology provides and anticipates how technology transforms the nature of practice. Probably, this would facilitate a system performance enhancement. Moreover, this is consistent with insights by Ackoff [11], who stated that in any system, when one improves the performance of the parts taken separately, the performance of the whole does not necessarily get improved because the way the parts fit together determines the performance of the system and not on how they perform taken separately. Thus, a system is not the sum of the behaviour of its parts, it is a product of their interactions. Further, Grabbe et al. [12] designate the frequent argumentation of AD as too simplistic thinking that falsely links the logic of a clear causal link. Rather, according to Bengler et al. [13], the human driver is both an active and passive participant in an accident, as well as accident avoidance and compensation element in the same system. Consequently, a more differentiated view is required to improve the safety of the entire road system through the efficient interaction between humans, machines, and other road users.

Therefore, according to Grabbe et al. [12], the very first step in the development process of automated vehicles is to understand the mechanisms of accident development and accident avoidance in road traffic. Hence, the driver's contribution in the corresponding situations can be assessed to derive requirements for automation and the potential of automation with its accompanying factors. Thus, the understanding of the mechanisms of the road system is essential. Otherwise, adverse automation surprises, as happened in other domains such as aviation, e.g., [14–17], will probably occur due to safety blind spots [18] and could ultimately develop into a "showstopper". This raises the question of: which methods are best suited to reveal the safety mechanisms in road traffic? To clarify the question, Grabbe et al. [12] reviewed the historical development of accident analysis models, the properties of the road system, and a common understanding of safety. In the following, a brief overview of their analysis and main conclusions is given.

According to Hollnagel [19], three different and major accident models can be distinguished: sequential, epidemiological, and systemic. Sequential accident models describe the accident as the result of a chain of discrete events occurring in a particular time sequence. Here, losses are caused by technical failures or human error, assuming that the cause-and-effect relationship is linear and deterministic [20]. Then, the focus changed with introducing the epidemiological models to an organisational level, where accidents result from a combination of different interacting factors [20]. This improved the understanding of accidents regarding complexity, but the causality is still linear and the links between states are loose, which that does not adequately represent the dynamics of a system [21]. Thus, systemic models arose seeing the accident process as a complex and interwoven event that cannot be broken down into its individual parts [22] and rather analysing interactions within the whole system. Salmon et al. [23] concluded that the road system, which connects technical, human, and social elements to transport people and goods from one place to another, is of a socio-technical nature. Additionally, they also argued in favour of a complex system based on the prerequisite properties of complexity presented by Dekker et al. [24]. Further, Perrow [25] defined a framework called interaction-coupling matrix to classify systems based on their system characteristics. Here, systems can be generally distinguished by the two dimensions of interaction and coupling. The interactions can be linear or complex and the couplings are loose or tight, which results in four quadrants of system assignments.

In particular, a socio-technical system (STS) can be described by an increasing number of tight couplings and complex and non-linear interactions [25]. Wienen et al. [22] extended this framework by combining it with the three types of accident models mentioned above. Based upon this, Grabbe et al. [12] classified the road system as a system with highly complex interactions and tight couplings assigning it to the systemic quadrant within the matrix. Consequently, they concluded systemic methods are best suited to represent a safety assessment of AD and especially reveal mechanisms in road traffic. This is also confirmed by the analyses of Larsson et al. [26] and Hughes et al. [27] claiming that system theory and systemic models are an important and major basis for safety work in road traffic. Furthermore, safety is a complex issue, and many different views exist, providing a variety of measures giving a reasonable description. So, every view captures some elements of safety but not the entire picture [28]. Thus, safety cannot be defined by one clear definition or construct. However, the historical development of the scientific study of safety points out two fundamental concepts where safety is concerned: the traditional thinking about safety, also called safety-I [29], which is based on the Newtonian and reductionist approach [30], and the modern view of risk and safety management, also called safety-II [29], which follows a complexity-oriented holistic approach [31] based on Resilience Engineering (RE) [32]. A critical perspective on the two safety perspectives concerning AD and road safety by Grabbe et al. [12,33] unveiled that the safety argumentation, as well as the safety assessment of AD, is largely safety-I driven, and the safety-II perspective is strongly neglected. Thus, they requested an urgent application of this safety view. This is also in line with Hollnagel's [34] (last slide) general statement regarding safety management that "it is an unavoidable dilemma that we inadvertently create the challenges of tomorrow by trying to solve the problems of today with the mindset (models, theories & methods) of yesterday".

Since systemic analysis methods, as well as a safety-II perspective, seem best suited to identifying the safety mechanisms in road traffic against the background of AD, Grabbe et al. [12] extensively compared the major systemic methods and discussed their benefits and limitations. The authors recommended using the functional resonance analysis method (FRAM) at the very beginning of a product development cycle, concluding that FRAM is the most adequate method to reveal the safety mechanisms in road traffic. This is also supported by Ferreira and Cañas [35], who see FRAM as a useful tool to build an understanding of the actual system mechanisms and workings that are needed to support the risk management concerning the proactive assessment of technological changes and their impacts. Apparently, there is no "one-size-fits-all" solution to safety, which is especially true for complex and dynamic STS. Thus, overall, we need combinations of different views, approaches, and measures including safety-I and safety-II. However, a significant perspective, that is RE, is currently lacking and inevitable as a fundamental basis for the safety assessment of AD. Here, FRAM, which is based on RE, is the most recent and promising step to understanding STS [36]. Therefore, Grabbe et al. [12] investigated the applicability of FRAM in a case study to evaluate its suitability in more detail with regard to a purely methodical process. They discussed several strengths and limitations. Ultimately, the authors concluded that the safety challenge of AD requires the study of interactions and mechanisms of the road system where FRAM adequately addresses these issues considering this method a "missing piece in the puzzle" for a risk assessment of AD, which potentially helps to reveal hidden risks or safety blind spots of AD.

To continue this research and to reduce the research gap in the safety assessment of automated vehicles regarding the aforementioned perspective, this paper aims to identify the mechanisms of road traffic in one specific scenario, that represents a huge potential to increase safety through automation in a complex setting, by using FRAM. Hence, these mechanisms can be compared between a human driver and a highly automated vehicle, which allows us to evaluate the contributions of the human driver and the automation. Finally, system design recommendations for AD, considering potential accompanying factors as well as insights for the validation process, reducing its effort, can be delivered in order to show how FRAM can be used for a systemic function allocation for the driving task between humans and automation.

The remainder of this paper is structured as follows. Section 2 summarises the theoretical foundations and individual analysing steps of FRAM, as well as its applications. Section 3 describes the implementation of the overall methodological research process and its individual steps in detail. In Section 4, the results are presented including the FRAM model of the analysed scenario, the identification of the contributions by the human driver and the automation to the safety of the system, and recommendations for system design as well as the validation process of AD. Then, Section 5 discusses the results with respect to the research goals and also outlines methodological issues. Finally, a brief conclusion and outlook for future research are given in Section 6.

2. Functional Resonance Analysis Method

FRAM [37] is basically a qualitative method for risk assessment and accident analysis. It allows the modelling of mechanisms within a complex STS, including their interfaces between humans and technology, coupling and dependency effects, nonlinear interactions between elements, and functional variability [38]. The purpose of the resulting model is to analyse how something happens or how a system works as work-as-done (WAD). In particular, the description and understanding of the STS are given in terms of functions rather than components. A FRAM model focuses on adjustments to everyday performance, which usually contribute to things going right. Rarely, these performance adjustments aggregate in unexpected ways, functional resonance will occur, and accidents are the most extreme result. The ultimate objective is not to eliminate performance variability but to investigate and monitor what is necessary for everyday performance to go right, trying to dampen variability in order to reduce resonance effects and unwanted outcomes [37]. In general, the results of a FRAM analysis contribute to the understanding of real work and unveil unsafe functional interactions within one agent and between different agents that are often underestimated by traditional methods and design approaches [35,39].

FRAM follows four principles (i.e., the equivalence of success and failures, approximate adjustments, emergence, and functional resonance), and four steps (i.e., modelling the system through identifying its functions, identifying the function's performance variability, aggregating the variability, and managing the variability) are required for its analysis as detailed in Hollnagel [37]. The steps are briefly described in the following. In the first step, the essential functions of the system ensuring the success of everyday work are identified to build a model. These functions produce a certain outcome referring to tasks as work-as-imagined (WAI) or activities as WAD. Each function is characterised by six aspects (i.e., input, output, precondition, resource, control, and time), which couple each function with several other functions representing a specific instantiation of the model. The resulting model is traditionally represented graphically by hexagons depicting each function with its six aspects. Furthermore, the functions can be divided into two classes: foreground and background functions. Foreground functions are the focus of the analysis and may vary significantly during an instantiation of the model. In contrast, background functions are stable and represent common conditions as system boundary that are relevant for and used by foreground functions. The second step is to identify and specify the performance variability of each function. This is crucial to understand how the variability can propagate through the system by the couplings between functions, which can lead to unwanted outcomes. After the identification process, the variability has to be characterised using different variability manifestations, the phenotypes. The simple solution considers two phenotypes, these are timing and precision, where the function's output in terms of timing can occur too early, on time, too late or not at all, and in terms of precision, the output can be precise, acceptable, or imprecise [37]. As it is not enough to simply know the variability of individual functions in isolation, the third step in FRAM is to aggregate the variability to know where functional resonance emerges. This is done by defining upstream-downstream couplings where variability can be caused through couplings of upstream functions, when the output used as, for example, input or precondition is variable and thus affects the variability of downstream functions. This impact is likely to lead to an increase in variability (amplifying effect), a decrease in variability (damping effect) and to maintain variability (no effect). The last and fourth step consists of the monitoring and management of the performance variability that was identified in the previous steps. This step aims to manage or dampen variability to a level where no unwanted outcomes arise, rather than eliminating variability since this is inevitable for things going well in complex STS. Finally, this ensures the safety and performance of the system. The implementation of each step is more detailed in Section 3.

In the past, FRAM has been widely used, applied, and enhanced methodologically in a variety of domains for retrospective as well as prospective analyses, as detailed in a comprehensive review by Patriarca et al. [40]. Hence, FRAM has been progressively evolved since its starting point in 2004. The main application fields include aviation, e.g., [41–43], healthcare, e.g., [44–46], industrial operations in plants, e.g., [47–49], the oil and gas industry, e.g., [50,51], and maritime, e.g., [39,52,53] and rail transport, e.g., [54,55]. However, the context of road safety has seldom been addressed by FRAM. Here, applications refer to road safety management in a case study in Myanmar [56], a comprehensive comparison of FRAM with other systemic methods regarding the safety mechanisms in road traffic, as well as a thorough investigation of FRAM's applicability in a case study evaluating its suitability with regard to a purely methodical way against the background of the impact of introduced automation [12], and a safety analysis of conditional automated driving including the human-machine collaboration in the event of an authority transfer from the automated system to the human driver in time-critical situations [57].

3. Research Method

3.1. Overall Methodology

As mentioned above, FRAM is a qualitative research method, which implies that classical statistical procedures applied to quantitative methods are not adequate to meet the three quality criteria in quantitative terms of internal and external validity, reliability, and objectivity. To overcome this issue, we applied the approach of Anfara et al. [58], translating the quality criteria in qualitative terms into credibility, transferability, dependability, and confirmability in order to better assess the research quality and rigour in this study and thus to improve their trustworthiness. Additionally, Creswell and Miller [59] identified several verification strategies to comply with the four qualitative terms, where Creswell and Poth [60] recommended that at least two of these strategies be used in any qualitative study. The assignment of the quality criteria in quantitative and qualitative terms, as well as their verification procedures, can be taken from Table 1. Here, the verification strategies underlined boldly are implemented in this study to fulfil the four qualitative terms.

Table 1. Assignment of the quality criteria in quantitative and qualitative terms as well as their verification procedures based on Anfara et al. [58] and Creswell and Miller [59].

Quantitative Term	Qualitative Term	Verification Strategies
Internal validity	Credibility	Prolonged engagement in field; Use of peer debriefing; Triangulation; Member checks; Time sampling; Persistent observation; Clarifying researcher bias
External validity	Transferability	Provide a thick description; Purposive sampling
Reliability	Dependability	Create an audit trail; Code-recode strategy; Triangulation; Peer examination; Stepwise replication
Objectivity	Confirmability	Triangulation; Practice reflexivity

As described in Section 2, the FRAM method comprises four main methodological steps. These steps and their underlying substeps are shown in Figure 1. The aforementioned quality criteria and verification strategies are intertwined in these steps. The following subchapters will explain the respective steps in detail.

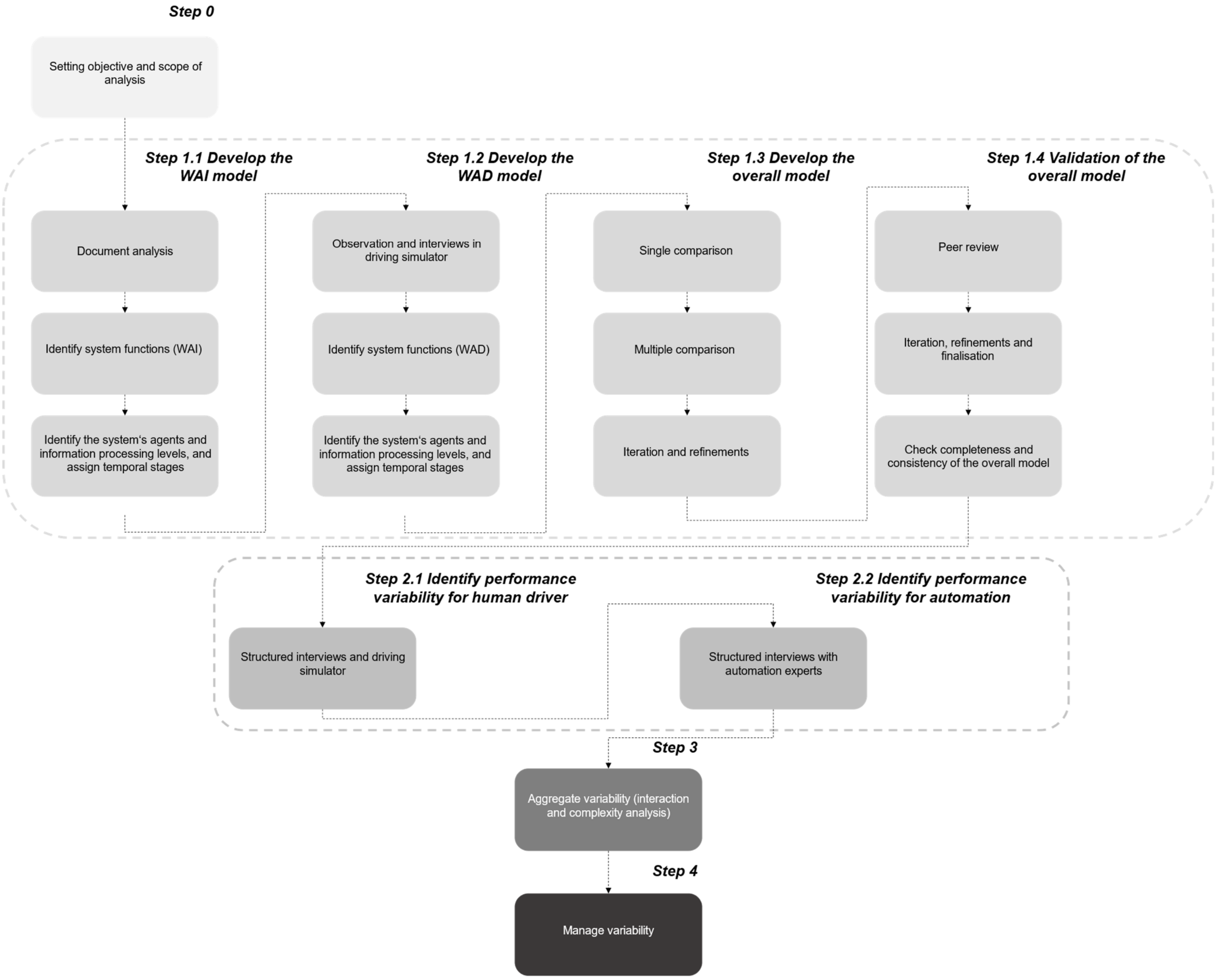

Step 0
Setting objective and scope of analysis
Step 1.1 Develop the WAI model
Document analysis
Identify system functions (WAI)
Identify the system's agents and information processing levels, and assign temporal stages
Step 1.2 Develop the WAD model
Observation and interviews in driving simulator
Identify system functions (WAD)
Identify the system's agents and information processing levels, and assign temporal stages
Step 1.3 Develop the overall model
Single comparison
Multiple comparison
Iteration and refinements
Step 1.4 Validation of the overall model
Peer review
Iteration, refinements and finalisation
Check completeness and consistency of the overall model
Step 2.1 Identify performance variability for human driver
Structured interviews and driving simulator
Step 2.2 Identify performance variability for automation
Structured interviews with automation experts
Step 3
Aggregate variability (interaction and complexity analysis)
Step 4
Manage variability

Figure 1. Methodological steps of the research method implemented in this study.

3.2. Step 0: Selection and Description of Scenario: Setting the Objective and Scope of Analysis

In this work, FRAM was used as a method for a qualitative/quantitative proactive risk assessment. Thus, the scope of analysis and the degree of resolution have to be described to set the scene and system boundary for the four steps that follow. In particular, a scenario-based analysis was conducted to compare the contributions between a human driver and AD and to evaluate their potential effects in order to improve the system design. The scenario is described below.

The scenario selected was an overtaking manoeuvre on a rural road. The main reasons are as follows. First, accidents in the city and on rural roads are by far the most critical, considering the accidents according to their location concerning frequency and severity in Germany. Furthermore, 58% of all fatal accidents in 2018 in Germany occurred on rural roads. Second, on rural roads, collisions with oncoming vehicles and leaving the carriageway pose the greatest danger [61]. By far the largest proportion of collisions with oncoming vehicles is caused by overtaking manoeuvres [62]. Therefore, overtaking situations represent accident black spots on rural roads, offering great potential for road safety improvement. Additionally, overtaking situations are classified as a relevant scenario category for a scenario-based validation of AD [33]. Third, according to Netzer [63], overtaking is a very complex traffic process with a variety of influencing factors involving several different subtasks, such as swerving, adjusting speed, merging, and the interaction of at least two drivers. Thus, this scenario offers a great potential to highlight the interaction and complexity of road traffic, including the systemic interdependencies between different road agents and the environment. In addition, results might be transferred to other road traffic scenarios because overtaking situations make up a large part of everyday driving tasks. Overall, the overtaking situation on rural roads is a good starting point for a socio-technical analysis under the lens of RE.

Figure 2 schematically depicts the overtaking scenario. This consists of four road users or agents: the ego vehicle (EV), the lead vehicle (LV), the rear vehicle (RV), and the oncoming vehicle (OV). Behind the OV, identified by the second orange and unlabelled vehicle, other vehicles form a line of cars. However, these vehicles and drivers are not considered agents for the modelling and scope of analysis and are therefore out of system boundary. To get a better overview, the scenario can be divided into five temporal and spatial stages from EV's point of view (see Figure 2): following a vehicle in front, swerving into the oncoming lane, passing the leading vehicle, merging back into the starting lane, and getting in the lane again.

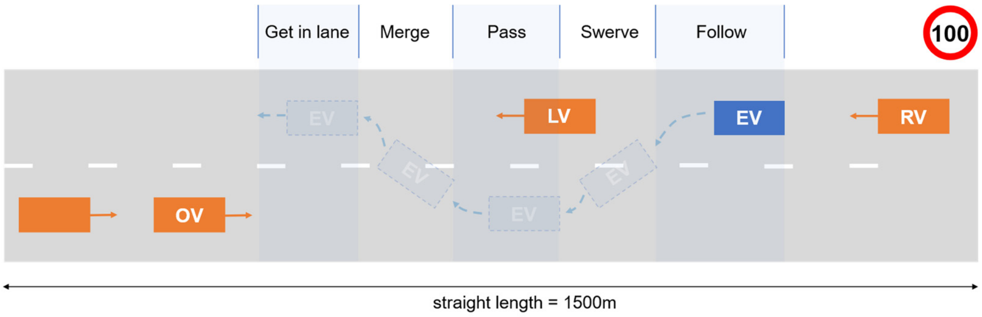


Figure 2. Schematic illustration of the overtaking scenario comprising different road users/agents and divided into five temporal and spatial stages. EV = ego vehicle, LV = lead vehicle, RV = rear vehicle, OV = oncoming vehicle.

The four agents are driving on a straight rural road for a distance of 1500 m with no vertical elevation, on which the maximum speed limit is 100 km/h, overtaking is permitted and no obstructions exist. One lane runs in each direction and the median is dashed. The total width of the road is 6 m. The road is well constructed and all necessary road markings are in place. On the side of the road, there is light vegetation. The weather conditions are sunny and dry.

The EV is following the LV and at the same time followed by RV. The LV is driving at a speed of 80 km/h. In the oncoming traffic, a vehicle OV and following vehicles are coming towards at 100 km/h with different time gaps. In principle, the OV represents the oncoming traffic. All agents always keep the necessary safety distance to their vehicle in front and comply with the traffic regulations. The EV is under time pressure and wants to reach its destination quickly, and since LV is travelling below the speed limit, it starts an overtaking manoeuvre. The other agents are reacting to the overtaking manoeuvre of EV. In general, the EV is driven once by a human driver and once by an automated system (SAE-level 4) according to SAE J3016 [3] with no car-to-x communication. The other vehicles are always driven by a human driver in both cases. Overall, the overtaking scenario should represent a simple and everyday overtaking manoeuvre on a rural road, in which four road users are interacting primarily with one other. This represents a scenario in which most overtaking accidents occur, that is a straight flat section in daylight and on a dry rural road, all in all, under good external conditions [62].

3.3. Step 1: Identification and Description of the System's Functions

3.3.1. Develop the WAI Model

The WAI model is based on a comprehensive and detailed hierarchical task analysis of driving developed by Walker et al. [64]. This work is created on a task analysis conducted by McKnight and Adams [65] in 1970, the UK Highway Code, several driving standards and manuals, input by subject matter experts (SMEs), and numerous on-road observation studies. The tasks and plans are constructed using logical operators such as And, Or, If, Then, Else, While, and so on. The list of tasks and plans, which are essential for the overtaking scenario, were translated into functions where the logical operators were used to define couplings between each function through their aspects. First, a WAI model was created for each agent, followed by a WAI model combining all agents in one model assigned to the five temporal stages of the scenario. In addition, the functions were labelled and distinguished by different information processing levels.

3.3.2. Develop the WAD Model

Since it is not sufficient to know only the theoretical mechanisms of the overtaking process, the next step is to create a WAD model using observations and interviews implemented in a driving simulator study which serves to update and enhance the WAI model into a more realistic overall model.

Driving Simulator

Here, a static driving simulator (see Figure 3) was used. The environment is simulated by three flat screens with a resolution of 4K covering the space from the left-side window to the right-side window of the car, which ensures a 120° viewpoint in front. Additionally, the rear-view mirror is virtually displayed at the top of the centre screen. The side mirrors are displayed via two small monitors placed to the left and right of the subject. The driver, seated on a default automobile seat that is adjustable in height and longitudinal direction, has a steering wheel for lateral control that can be adjusted along the axis, as well as an accelerator and brake pedal for longitudinal control. The use of a turn signal and a shoulder view to the rear are not possible. Behind the steering wheel is a combination display that shows the engine speed and the current speed of the vehicle. Further, the driving simulator is equipped with automatic transmission and sound, consisting of engine, environmental, and vehicle noises that are reproduced via two speakers placed next to the pedals. During a test drive, the room was darkened to increase the immersion for the driver. SILAB 6.0 of the Würzburg Institute for Traffic Sciences GmbH in Germany was used as the simulation software.

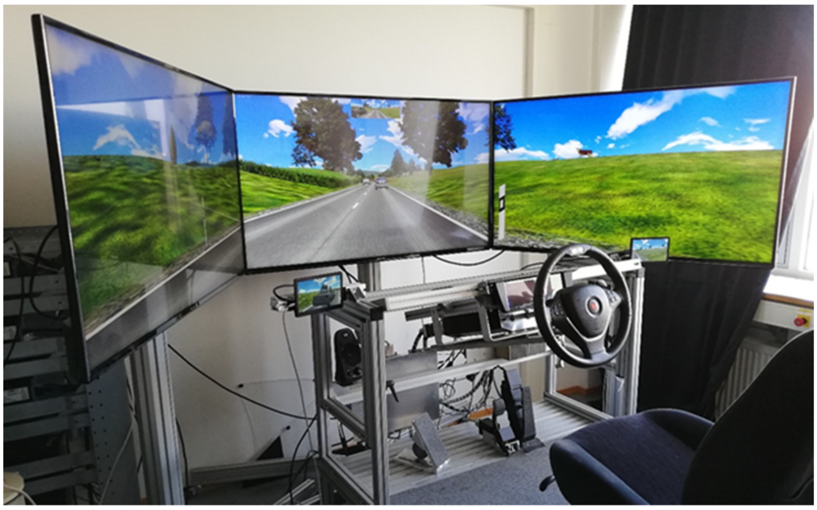


Figure 3. Structure of the static driving simulator.

Sample

A total of 10 participants took part in the study. Of these, seven were men and three were women with an average age of 28 years (*SD* = 2.26 years), ranging from 24–31 years. All owned a valid driving licence and drive an average of 18,000 km a year

(*SD* = 10,055 km/year), which shows a solid experience in road traffic. Furthermore, all subjects have already participated in a driving simulator test and were well acquainted with the driving simulator, which is why it can be assumed that their real driving behaviour has not changed much in the driving simulator. This is consistent with the indication that 80% would perform similar driving manoeuvres and overtaking manoeuvres in reality. The driving styles were heterogeneous, ranging from safe and leisurely to slightly risky and fast-paced, which was surveyed using a 5-point Likert scale.

Procedure

First, the subjects were informed about the goals and content of the study and signed an informed consent. Afterwards, the subjects took a seven-minute test drive, which included everyday driving scenarios on rural roads, to learn about steering, braking, and the driving simulator system. Then the actual test drive began. Here, the driving data, as well as the audio track and the subject's behaviour, were recorded for evaluation. In total, the experiment lasted 30 min, and each subject experienced the scenario from the perspective of each of the four agents, in which the order of perspectives was as follows: EV, LV, RV, OV. The subject passed through each perspective three times. The first pass of the overtaking manoeuvre was used for familiarisation, during the second the subjects were asked to think aloud and explain their actions over the following few seconds, and during the third pass, the simulation was stopped five times (which represented the five stages of the scenario, see Figure 2) whereupon the subjects were asked to explain in detail which functions they would perform over the next few seconds. The functions refer to the three information processing levels of perception, cognition, and action. Between the actual test scenarios, that is the overtaking manoeuvre on the straight rural road, the test subjects each drove a small winding course through a wooded area so that the entire scenario would appear as natural as possible. After the test drive, subjects completed a short questionnaire to collect demographic data. Additionally, driver type data, as well as perceptions in the driving simulator test, were surveyed. Finally, a semi-structured interview was conducted. The interview queried specific aspects of the overtaking process from the perspective of all four agents that had not been considered before. The interview consisted of ten questions. The first six questions related to the execution of the overtaking manoeuvre regarding the five stages. The subject described, for example, the information on which their decision to start an overtaking manoeuvre was based, as well as its concrete execution. In addition, it was asked how the driver determines whether a current overtaking manoeuvre is at risk, how he/she reacts, and how a manoeuvre is successfully completed. The last four questions were general in nature (e.g., perception of environmental influences, the influence due to time pressure, or factors that can trigger a critical situation).

Measures and Analysis

In the evaluation to identify and describe the system's functions, the interviews, as well as the audio track and the driving and behavioural driver data, were used. The responses in the interviews, as well as the audio track during the experiment, were collected, categorised, and assigned frequencies. From this processed interview data, as well as the objective data streams such as the longitudinal and lateral driving behaviour in response to scenario objects or the behaviour of other drivers, activities for driving tasks were identified and subsequently translated into functions. This finally led to the WAD model, where the individual functions were linked based on the observations.

3.3.3. Develop the Overall Model

As a first step, each of the two researchers compared the WAI and WAD models they had created individually and tried to unify them into an overall model. The procedure was such that the WAI model formed the basis and newly discovered functions and couplings were added by the WAD model. After this, the two individually generated overall models were combined using a joint comparison and discussion by the two researchers. In a final step, the researchers refined the complete overall model in iterative steps by going through the model using an in-depth cognitive walkthrough to recognise potential missing functions or couplings and falsely linked functions. The overall model, as well as the WAI and WAD models, were produced using the software FRAM Model Visualiser (FMV) [66] Pro 2.1, available at http://www.zerprize.co.nz/FRAM/index.html (accessed on 25 August 2021).

3.3.4. Validate the Overall Model

In the last step, the overall model was calibrated and validated through a focus group within a peer review workshop to ensure objective, reliable, and valid analysis results based on the FRAM model. The peers were seven experts (5 male, 2 female) with strong knowledge and broad experience of human factors in the automotive area. The experts were educated about the FRAM model and its creation process one week before the workshop through a 90-min recorded video. In addition, general background information about FRAM was given to familiarise the peers with the method, and participants were divided into three groups (EV; LV; RV & OV) to provide comments on the specific agents. In the workshop, the overall model was then discussed step by step for each agent. However, it turned out that the planned format was inefficient. Therefore, in three separate two-and-a-half-hour meetings, the model was explained and discussed again in detail for the respective three groups, and the experts then gave their feedback and the models were iteratively adapted. At a follow-up meeting, the overall model was finally iteratively calibrated and fine-tuned again with all seven peers in a joint two-hour session. To validate the overall model, the peer group reflected on their personal experience and human factors knowledge of driving a car, including manual driving as well as automated driving. This contained additions, modifications, or deletions regarding functions and their couplings, as well as the assignment of agents, temporal stages, and information processing levels. Having agreed that the overall model accurately reflects the essential mechanisms of the overtaking scenario, the last step was a formal validation. Here, the model has been checked and adjusted for consistency and completeness, using another software facility, the FRAM Model Interpreter [66,67], which is incorporated into the FMV Pro. It was a stepwise automatic interpretation of the syntactical and logical correctness of the overall model.

3.4. Step 2: Identification of Performance Variability

3.4.1. Identify Performance Variability for the Human Driver

The identification of the performance variability for the human driver was twofold and was based on objective as well as subjective data, as described below.

Driving Simulator Study

First, a second driving simulator study was conducted. The simulator environment and the setting were the same as mentioned in Section 3.3.2.

Sample

Overall, 30 subjects (20 males, 10 females) including German students and scientific employees, aged between 21–30 years (*M* = 24.84 years; *SD* = 2.96 years), took part in the study. All had a valid driving licence and drive an average of 11,724 km a year (*SD* = 7742 km/year). Furthermore, half of all subjects had already participated in a driving simulator test. Additionally, 80% would perform similar driving manoeuvres and overtaking manoeuvres in reality. All subjects had experienced driving skills, with 76% driving daily to weekly. The driving styles were heterogeneous, ranging from safe and leisurely to slightly risky and fast-paced.

Procedure

Overall, the experimental track was the same as mentioned in Section 3.3.2. Before the test drive, the subjects were informed about the goals and content of the study and signed an informed consent. Afterwards, they took a 15-min test drive on a rural road for familiarisation. According to the Wiener driving test [68], an observation period of about 15 min is necessary before drivers show their everyday normal driving behaviour and fall into their regular habits, which should ensure a valid investigation of everyday performance variability. Then the actual test drive began. Besides the recording of driving data, audio track, and the subject's behaviour, the glance behaviour was tracked with a head-mounted eye-tracking system via Dikablis Glasses 3 from Ergoneers in Germany. This ensured insights, especially into the drivers' perceptual behaviour, in addition to executive activities, and to record cognitive processes. The participants drove the four agent perspectives three times in permutated order, intending to reproduce their everyday driving behaviour and complete overtaking manoeuvres and driving tasks as quickly as possible, but as safely as necessary.

Measures and Analysis

To determine performance variability, the driving data and glance behaviour were evaluated for each run (a total of 90 data sets per agent and function), with each run then assigned to the different characteristics of the timing and precision phenotypes based on previously established definitions of the characteristics of the phenotypes per function. Here, Table 2 exemplifies this for the lane-keeping function.

Table 2. Definition of the timing and precision characteristics using the lane-keeping function as an example.

Phenotype	Characteristic	Definition
Timing	Too early	If the driver already countersteers although the vehicle is driving in the middle of the lane.
	On time	If the driver countersteers in time (the vehicle is approaching the left or right of the lane boundary) to keep the vehicle in the lane.
	Too late	If the driver countersteers too late (vehicle has already left the lane) to keep the vehicle in the lane.
	Not at all	If the driver does not countersteer at all to keep the vehicle in the lane.
Precision	Precise	If the car always drives perfectly along the centre line between the left or right of the lane boundary.
	Acceptable	If the car always drives between the left or right of the lane boundary.
	Imprecise	If the car crosses the left or right of the lane boundary.

Finally, this resulted in a frequency distribution of performance variability for each function as an average over all runs (e.g., for timing 90% on time and 10% too late and precision 20% precisely and 80% acceptably). The reason for specifying performance variability via a frequency distribution is to create as realistic as possible a representation of actual everyday performance.

Interviews and Survey

Unfortunately, only a few functions' performance variabilities (mainly functions referring to actions) could be objectively and reliably determined by observation in the driving simulator, and a large part of the perceptual and cognitive processes could not be assessed. Thus, large-scale structured interviews combined with a survey were conducted in a second step. In general, the following rule applied to determine the variability of performance per function: If the variability of a function could be objectively recorded in the simulator study, then these values were used, if not, then the values from the interviews were used. Since most of the functional variability could only be captured subjectively through the interviews, the drivers' self-assessment had a primary role.

Sample

Overall, 30 subjects, who are a mixture of students, scientific employees, and people with completely different educational and occupational backgrounds from Germany, took part in the interviews. The participants (21 male; 9 female) have an average age of 32.33 years (*SD* = 12.35 years), with an age range of 21–61 years. All owned a valid driving licence and drive an average of 17,166 km a year (*SD* = 8971 km/year). All subjects had experienced driving skills, with 83% driving daily to weekly. Their driving styles were heterogeneous, ranging from safe and leisurely to slightly risky and fast-paced.

Structure of Questionnaire and Analysis

Because of the high number of functions, two questionnaires were created using the online survey tool LimeSurvey. They cover 100 functions and were gone through step by step in an interview so that queries could be clarified. The first questionnaire determined all driving tasks of LV, RV, and OV, the second one determined the variability for driving tasks performed only by EV, with each questionnaire being completed by 15 participants. Both questionnaires were already reduced by redundant functions, which means functions that are executed several times, that are in different stages, or by several agents. The structure of the questions is described in the following, which was inspired by the approach of Patriarca et al. [45], who conducted the determination of performance variability in a neuro-surgery healthcare setting via an online survey. The driving tasks were always queried according to the stages of the scenario and the subjects were informed of the stage in which the driving task was performed. For each driving task, the name of the driving task, which agent performs it, a description of the task of the function, and the output of the same were given. This was followed by the evaluation of variability in timing and precision. Here, the subjects stated in per cent how often they perform a driving task in everyday life: too early, on time, too late, or not at all. For this purpose, each of the sliders was moved in five per cent increments. For better orientation, value ranges were defined for the frequency categories: never (0%), rarely (1–25%), sometimes (26–50%), often (51–75%), usually (76–99%) and always (100%). The evaluation of precision was carried out in the same way, except that here the subjects indicated how precisely they perform the driving task in everyday life: precisely, acceptably, or unacceptably. The sum of the individual responses had to add up to 100 per cent in each case. Finally, the performance variability distribution ratings for each function were averaged for each characteristic over all participants.

Procedure

The procedure of the interview and the structure of the questionnaires were as follows. The subjects are first informed about the theme and procedure of the study and signed an informed consent. The interview lasted about 60 min. After that, the scenario, agents, stages, and structure of the questionnaire were explained. This was followed by a demographic questionnaire and a test question so that the subjects could familiarise themselves with the structure of the questions. Before the actual survey began, the subjects watched a video that visualised the scenario in real-time. During the survey, questions could be asked to eliminate misunderstandings.

3.4.2. Identify Performance Variability for Automation

Due to a lack of public data on AD performance and driving behaviour, structured interviews combined with a survey were also conducted to determine performance variability for automation as a generic concept based on the current state-of-the-art of automation systems and short-term developments.

Sample

Here, twelve experts (10 male, 2 female) participated in the interviews. Most of the experts came from suppliers or original equipment manufacturers (OEMs) in the German automotive industry, a few from German universities, and one from an OEM in the USA. The experts held various positions within the development of automated driving functions and had extensive practical and theoretical knowledge regarding the performance of current series and prototype functions. On average, the experts had been working in their current function for 5.83 years (*SD* = 5.34 years) and had already gained experience in the field of driver assistance or vehicle automation for an average of 8.33 years (*SD* = 4.79 years). Seven described their general attitude towards vehicle automation as consistently positive, four as positive but with reserved euphoria because of a clear necessary increase in reliability, and one was ambivalent, especially about implementing higher levels with broad application areas.

Procedure and Analysis

The procedure of the interviews as well as the structure of the questionnaires were the same as for the case of the human driver, as mentioned above. The functions for the EV were split up into two questionnaires due to the high number, each of which covered 41 driving tasks or functions. Each survey was completed by six experts. The only difference in the individual questions was that no frequency distribution concerning the characteristics for timing and precision had to be given, but only one characteristic per phenotype (single choice) was to be selected. That was considered the most probable in the analysed scenario for AD against the background of short-term automation developments. All ratings of every expert were then combined into a frequency distribution of performance variability for each function.

3.5. Step 3: Aggregation of Variability

The purpose of the third step is to look at how the variability of the functions aggregate and propagate through the system in a specific instantiation of the model to determine potential functional resonance leading to unexpected outcomes arising through interaction and complexity in the system. Because of the complex scenario and the fact that its qualitative modelling by FRAM was quickly becoming overwhelming, we enhanced the research by a semi-quantitative approach according to Patriarca et al. [69] and Grabbe et al. [12]. This was implemented with the help of the software myFRAM 1.0.4 [70], which was developed in Visual Basic for Applications and interfaced with Microsoft Excel and FMV, enabling the FRAM model to be converted into a matrix so that a quantitative or numerical calculation is possible. The structure of the defined metrics is shown in Figure 4. Here, the nodes represent the respective metrics, and the structure, that is which metrics are composed how, is marked by arrows and their direction from right to left. The green nodes will later be used as the main analysis metrics in Section 4.2. In general, the metrics can be divided into three categories: functional variability, system resonance, and system propagational variability. The functional variability represents the variability that a function directly receives and transfers without considering their interaction and effect in the system sufficiently. Therefore, the system resonance tries to reflect the interaction and complexity of a function in the system, incorporating non-linearity, emergence, and dynamic of the system. It is a kind of weighting of the impact and affectedness of a function to evaluate the effect of a function variability system-wide. Combining functional variability and system resonance results in system propagational variability, which shows the systemwide impact and affectedness of each function's variability up to a global system variability level. The definition and calculation of each metric within the three categories, which were implemented with myFRAM and MATLAB 2020, are described below.

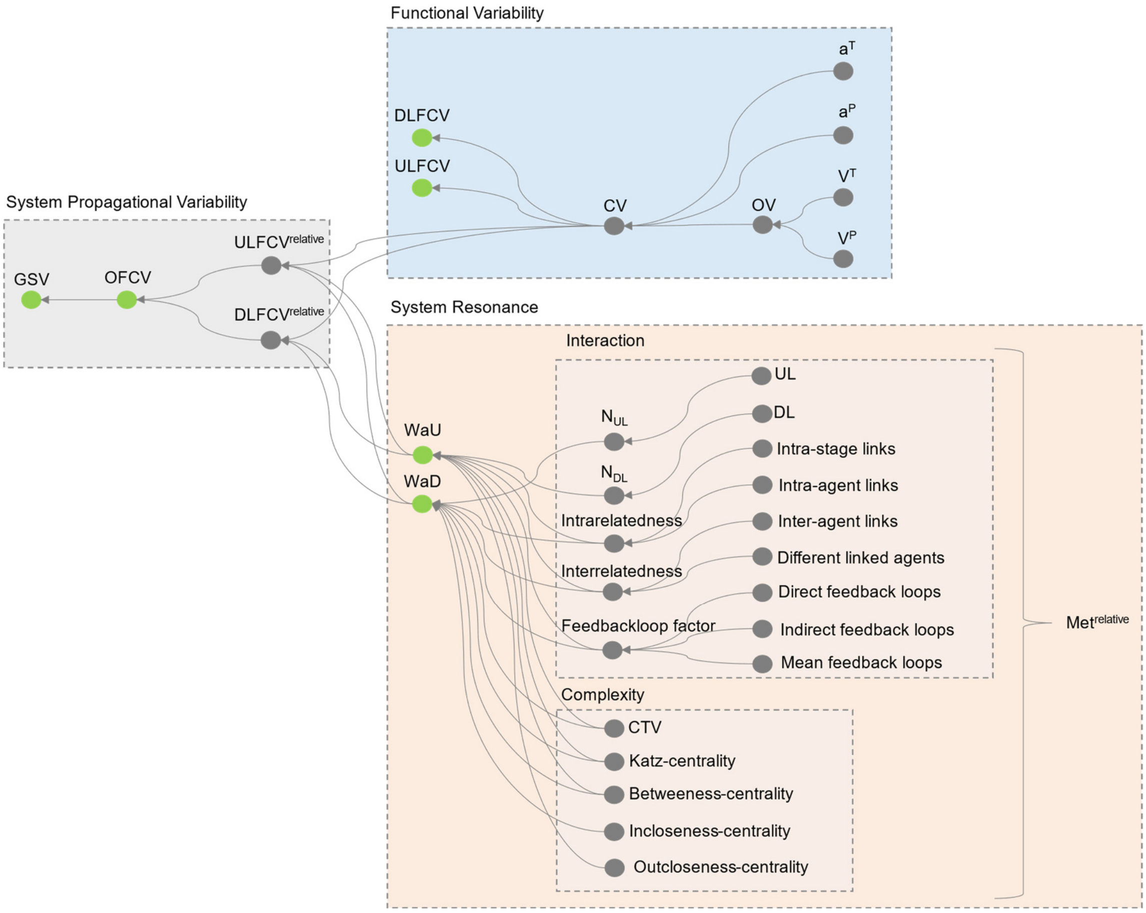


Figure 4. Structure of metrics by the semi-quantitative approach. The nodes in green represent the main analysis metrics.

3.5.1. Metrics for Functional Variability

The final calculation of functional variability is based on the downlink (*DL*) and uplink (*UL*) coupling variability (*CV*) of one foreground function (downlink functional coupling variability *DLFCV* and uplink functional coupling variability *ULFCV*). The *DLFCV* was used to understand the implications of the coupling variabilities of one entire upstream function *j* to associated downstream functions *i* and the *ULFCV* was used to comprehend the impact of the variability of a downstream function *i* through its incoming coupling variabilities of upstream functions *j*. The calculation formula for *DLFCV* and *ULFCV* can be seen in (1) and (2), respectively:

$$DLFCV_j = \sum_{i=1}^{j} CV_{ij} \tag{1}$$

$$ULFCV_i \; = \sum_{i=1}^{j} CV_{ij} \tag{2}$$

To keep the paper readable, the formulas of the remaining metrics on which the *DLFCV* and *ULFCV* are based can be found in Appendix A.

3.5.2. Metrics for System Resonance

The performance of the overall system, in this case the FRAM model, is more than the sum of its function’s variabilities, and rather is determined by the interaction and fit of the individual subsystems (within and between agents as well as between agents and the environment). However, the metrics mentioned above did not adequately represent this and are only considered as taken separately without interactions (except for the variability propagation factors). Therefore, we further defined several metrics, categorised into an interaction and complexity dimension, which should represent this inherent complexity, which incorporates non-linearity, emergence, and dynamic of the system. On the one hand, the connectivity/interaction of functions was determined with the following metrics in order to calculate the degree to which a function interacts with other functions or agents in the system:

- *Number of downlinks and uplinks (N_{DL} and N_{UL})* which show how many functions a function can directly influence and how many functions it is directly influenced by, respectively.
- *Intrarelatedness* expresses how many functions a function is linked to within an agent (e.g., EV) and within the same stage (e.g., Follow) or in different stages (e.g., Follow and Pass).
- *Interrelatedness* presents how many functions of other agents (e.g., LV and OV) a function is linked to and weights it with the number of different agents.
- *Feedback loop factor* reflects the extent to which a function's output can influence its input through direct and indirect feedback loops.

On the other hand, centrality measures from graph theory were used to represent the complexity of the system. The reason for this choice is that graph theory proved to be well suited to investigate some emergent non-linear characteristics of systems to express by other approaches and their used metrics have been already proven to succeed in explaining many features of complexity [71]. The translation of a FRAM model into a network by graph theory was already applied by Bellini et al. [72] and Falegnami et al. [71], showing general good integrability of these approaches to prioritise key functions in a FRAM model adopting centrality measures in order to reflect a combination of couplings' weights and connectivity. However, the studies also implied that several centrality indices, representing the importance of a node/function, exist and that it is difficult for a centrality measure to be considered the most representative of FRAM characteristics since peripheral nodes/functions can also be important. Thus, the most appropriate centrality measures should be identified on a case-by-case basis [73]. Therefore, the authors of this paper chose a mix of the following three different centrality indices and one own defined metric, assuming this would be the best way to represent this complexity:

- *Katz*-centrality depicts the relative degree of influence of a function within the system, showing the extent of indirect impact.
- *Incloseness*- and *Outcloseness-centrality* measure how central a function is located in a system and thus the more central a function is, the closer it is to all other functions and therefore has a high potential for functional resonance.
- *Betweenness-centrality* shows the degree of a function to bridge functions with other functions, which makes it a critical function for system success.
- *Clustered Variability (CTV)* shows how much upstream and downstream variability accumulates around a function to depict where groups of functions with high variabilities exist that are directly coupled.

To keep the paper readable, the formulas of the metrics for the interaction and complexity dimensions can be found in Appendix B. Below the calculation and meaning of the two main indicators of system resonance, the Weight as Upstream (*WaU*) and Weight as Downstream (*WaD*) of a function *f*, are explained. The *WaU* and *WaD* reflect the system effect of a function as an upstream and downstream function, respectively. This should simulate the interaction and fit between functions and their inherent complex interdependencies. The respective metrics are included in the calculation in a weighted manner. The assignment of these weighting factors with numerical values was subjective and is reflected in Table 3. The assignment follows the logic that some metrics weigh more heavily than others. For example, *interrelatedness* weighs more heavily than *intrarelatedness*, since this considers that influencing other agents has a higher system effect than only influencing one's own agent. The *WaU* and *WaD* are determined as follows (3) and (4):

$$
\begin{aligned}
WaU_f \quad &= \beta_1 * N_{DL}^{relative}{}_f + \beta_2 * Intrarelatedness_f^{relative} + \ \beta_3 * Interrelatedness_f^{relative} + \beta_4 \\
&*FeedackLoopFactor_f^{relative} \ + \beta_5 * \ CTV_f^{relative} + \beta_6 * Katz - centrality_f^{relative} + \beta_7 \\
&* Outcloseness - centrality_f^{relative} + \beta_8 * Betweenness - centrality_f^{relative}
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
WaD_f \quad &= \beta_1 * N_{UL}^{relative}{}_f + \beta_2 * Intrarelatedness_f^{relative} + \ \beta_3 * Interrelatedness_f^{relative} + \beta_4 \\
&*FeedackLoopFactor_f^{relative} \ + \beta_5 * \ CTV_f^{relative} + \beta_6 * Katz - centrality_f^{relative} + \beta_7 \\
&* Incloseness - centrality_f^{relative} + \beta_8 * Betweenness - centrality_f^{relative}
\end{aligned} \tag{4}
$$

Table 3. Allocation of numerical values of the weighting factors for the calculation of *WaU* and *WaD*.

Weighting Factor	Numerical Score
$\beta_1\ (N_{DL}/N_{UL})$	4
$\beta_2\ (Intrarelatedness)$	2
$\beta_3\ (Interrelatedness)$	2.5
$\beta_4\ (FeedbackLoopFactor)$	1
$\beta_5\ (CTV)$	1
$\beta_6\ (Katz - centrality)$	4
$\beta_7\ (In - /Outclosenness - centrality)$	2.5
$\beta_8\ (Betweeness - centrality)$	2.5

3.5.3. Metrics for System Propagational Variability

In the final step, the *WaU* and *WaD* are offset against the *CV* values of each function, resulting in a relative *DLFCV* (5) and relative *ULFCV* (6) considering the interaction of one function's down- and uplink coupling variability within the whole system, showing how a function affects the system and is affected by the system, respectively:

$$DLFCV_j^{relative} = \sum_{i=1}^{j} CV_{ij} * WaU_j * WaD_i \tag{5}$$

$$ULFCV_i^{relative} = \sum_{j=1}^{i} CV_{ij} * WaU_j * WaD_i \tag{6}$$

Finally, the overall functional coupling variability (*OFCV*) of a function *f* could be determined from this (7):

$$OFCV_f = ULFCV_i^{relative} + DLFCV_j^{relative} \tag{7}$$

This metric identifies critical functions with high potential for functional resonance offering functional prioritisation of their impact into the system in that, for example, a high value means that the function has a large systemic effect and/or is largely systemically affected and/or a high variability accumulates in and around the function.

In the last step, a global system variability (*GSV*) could be calculated to show the accumulated variability of all functions and their interactions of the whole system for one specific condition. This enables, for example, a comparison of system performance between a system where purely human drivers operate and one where an automated system operates with human drivers. The *GSV* is the sum of the *OFCVs* of *n* functions within the whole system (8):

$$GSV = \sum_{f=1}^{n} OFCV_f \tag{8}$$

3.6. Step 4: Management of Variability

The final step proposes ways to manage performance variability, especially possible conditions of functional resonance, that have been found by the preceding steps. In this work, we proceeded as follows. In general, we are aiming to improve the performance variability of the entire system for the given scenario by deriving system design recommendations through a well-reasoned function allocation, which will be shown in Section 4.3. To achieve this, the performance variability of the entire system is analysed by comparing the contributions between human driver and automation to road safety based on systemic mechanisms on both an abstract global level (see Sections 4.2.1 and 4.2.2) and a fine grain level regarding the individual functions (see Sections 4.2.3 and 4.2.4).

4. Results

In this section, the results are presented. First, the resulting overall FRAM model is described. Further, critical functions are identified and analysed in-depth to compare the positive and negative contributions of the human driver and automation to system behaviour. Finally, recommendations for system design as well as the validation process of AD are derived.

4.1. The Overall FRAM Model

The overall model comprises 285 functions (210 foreground functions (hexagons) and 75 background functions (rectangles)) with 799 couplings and is shown graphically in Figure 5. All functions within an agent exist only once and are then executed several times by other functions at different stages of the manoeuvre. The functions are assigned respectively to the four different agents (EV, LV, RV, and OV) and five temporal stages during the scenario (Follow, Swerve, Pass, Merge and Get in lane). This is a modification of the Abstraction/Agency framework by Patriarca et al. [74] into a Space-Time/Agency framework, which should ensure enhanced knowledge representation combined with a multi-dimensional approach that is two dimensions: the temporal-spatial levels and the agency levels. Since it is not effective to analyse an STS according to only one level [74], this approach makes it easier to with complexity that requires a system to be structured following different levels of analysis with different resolutions and perspectives [75]. This is shown by the interactions within an agent and between different agents at different temporal and spatial occurrences. The stages always refer to the perspective of the EV, which is the focus of analysis. The functions can only be executed within the assigned agent and the assigned temporal stage(s) but can be coupled with functions of all other agents and stages.

To make the model clearer, the functions have also been colour-coded according to the following pattern to specify the type of functions in more detail:

- Driving functions:
    - ○ Yellow → perception driving tasks (e.g., to monitor road layout ahead of LV)
    - ○ Blue → cognition driving tasks (e.g., to assess the opportunity to overtake safely)
    - ○ Green → action driving tasks (e.g., to decrease speed)
    - ○ Orange → main manoeuvre tasks (e.g., to follow LV)
- Functions affecting driving:
    - ○ Red → characteristics of the infrastructure (e.g., to provide road signs)
    - ○ White → characteristics of the environment (e.g., to enable clear view on the road ahead (weather conditions, etc.)
    - ○ Grey → technical functions of the vehicle (e.g., to provide steering wheel)
    - ○ Purple → information by the policy (e.g., to provide safe braking distances by Highway Code)

The driving functions are classified into three levels of information processing (i.e., perception, cognition, and action) adopting the framework of types and levels of automation regarding the four-stage model of human information processing provided by Parasuraman et al. [76]. This facilitates function allocation between humans and automation, that is, the design decision of which system functions are to be performed by humans and which should be automated and to what extent to improve system safety. Thereby, main manoeuvre functions bundle several driving functions, which are intended to improve clarity.

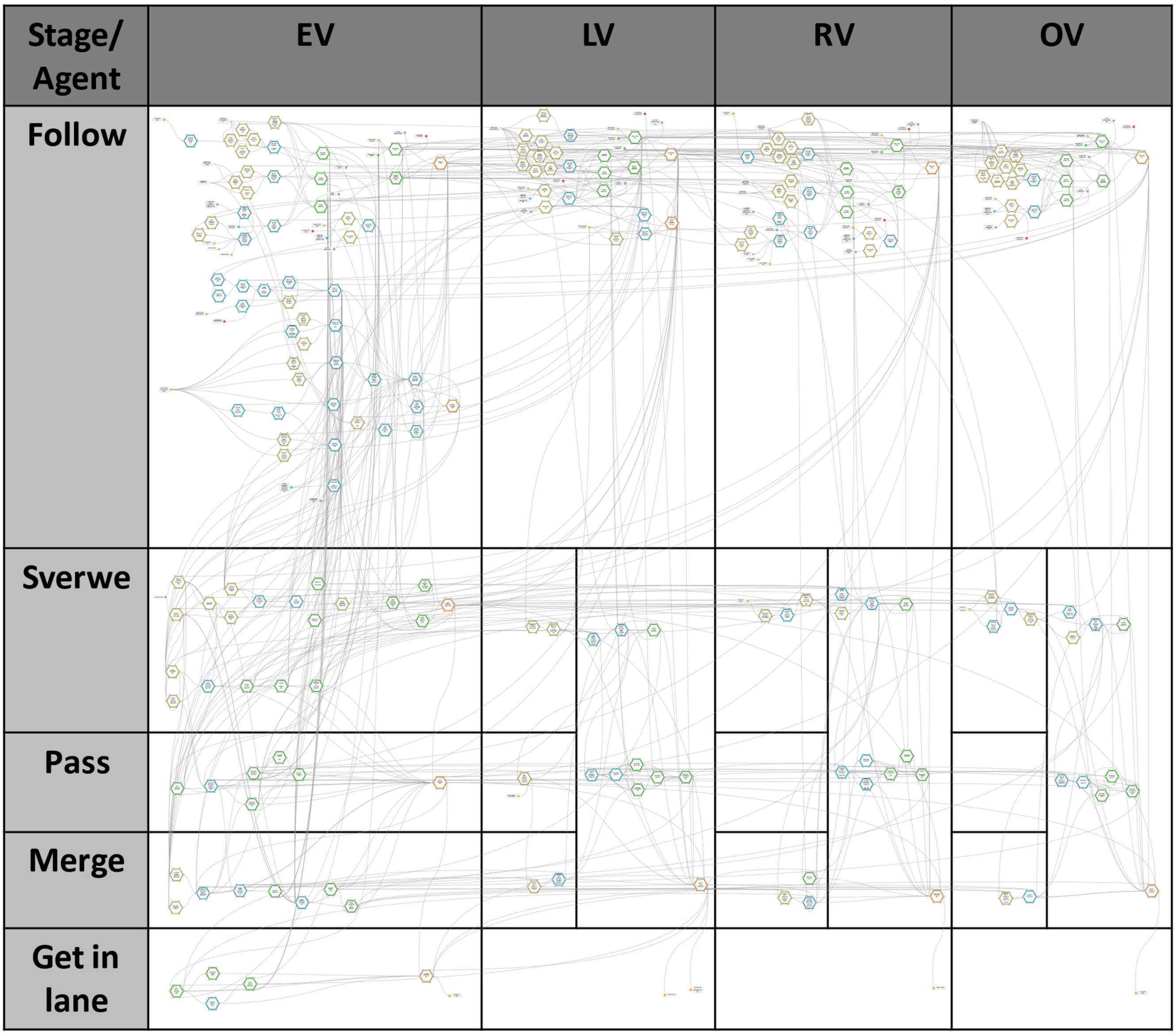


Figure 5. The overall FRAM model assigned in the Space-Time/Agency framework.

It should be noted that the model is the same for the human driver or the automation because of the assumption that there is no change in the functions of the system that have to be accomplished by the human driver or the automation. This is ensured by an appropriate resolution or abstraction of the functions. The difference between the two agents is only the variable performance of each function. The reason is that a FRAM model should treat humans and automation systems as equivalent producers of functions to compare the joint performance of both systems as the net result of the functional resonances as depicted by the *GSV*.

Due to the complexity of the model, we cannot represent and describe the actual structure and content of the whole model (the entire model can be viewed as an FMV data file in the Supplementary Materials S1). Therefore, we roughly describe the major functions per each agent and stage represented by the main manoeuvre functions in Appendix C in Table A3. Additionally, the driving behaviour of to follow by EV in the Follow stage (see Appendix C in Figure A1) is explained in detail to improve the comprehension of the remaining parts of the model.

4.2. Comparison of the Contributions between Human Driver and Automation to Road Safety Based on Systemic Mechanisms

In this subsection, the analysis process follows the hierarchical structure of the metrics depicted in Figure 4, moving from the abstract (left) to the detailed (right) focussing primarily on the main analysis metrics (green nodes). First, the abstract global analysis is accomplished through prioritising risk functions and analysing them in comparison across stages and function types between human driver and automation. Additionally, the global system variability is investigated. Second, the individual functional analysis is represented by distinguishing the interaction and variability of system functions to identify potential critical functional resonance, but also success factors, and finally analysing critical paths and their interactions in the system.

In general, a comparison of all system functions cannot be presented, so the following is an analysis of essential functions serving as examples to assist with comprehension of the derivation of system design recommendations in Section 4.3.

4.2.1. Prioritisation and Analysis of Risk Functions

The risk functions for human drivers and automation were identified through the analysis of the *OFCV* since this metric shows the criticality of a function measured by the system-wide impact of the function's variability. Here, the *OFCV* of each function was prioritised and ranked using the scree test (see Figure 6) according to Falegnami et al. [71]. Usually, the first knee is chosen to prioritise functions that lie left to the curve knee (that in our case filters only five functions, which are largely more critical than the following ones). However, as we are interested in focusing on a larger portion of risk functions, we needed a tool to help us decide which curve knee to use. Thus, we enhanced the scree test by a regression line. The rightmost curve knee, which lies above the first intersection point of the regression line (i.e., functions that lie above the average linear slope and thus differ significantly from functions below the average linear slope), is ultimately used as the decision criterion. Thus, we selected the third knee, allowing us to consider 23 risk functions for the human driver. The selection process for the automation was the same, resulting in 22 risk functions. A list of risk functions is shown in Appendix D in Table A4. The risk functions are not only related to the agent EV, but also the other agents. Considering a function allocation for the system design (which will be explained in more detail in Section 4.3.1), the following should be taken into account. If a function is only an automation risk, it is recommended that it should be performed by humans, and vice versa. However, if a function poses a risk to both, it is necessary to analyse thoroughly which control mode seems to be the best.

As seen in Figure 7, the most risk functions are in the Follow stage, which also includes significantly more functions, however. In the other stages, the distribution is about the same, except for the Swerve/pass/merge stage, in which humans have six times more risk functions than automation. However, the risk functions in this stage are all performed by other agents than EV, so it can be interpreted that the other agents are more negatively influenced by the human driver of EV than by the automation. However, this would need to be verified since the other agents are only influenced by action functions and these are predominantly performed worse by the human. Furthermore, the data from other agents are only based on experiences with human drivers and not with automation. Moreover, the Get in lane stage is the only stage without a risk function.

Figure 8 shows that the risk functions for automation are mainly loaded by perception and cognition. Merely one third relates to action and main manoeuvre functions. In humans, on the other hand, mainly action functions and the main manoeuvre functions are considered risk functions, whereby the main manoeuvre functions are predominantly action-intensive. Only one fifth is accounted for by cognition functions, and perceptual functions do not pose any risks at all.

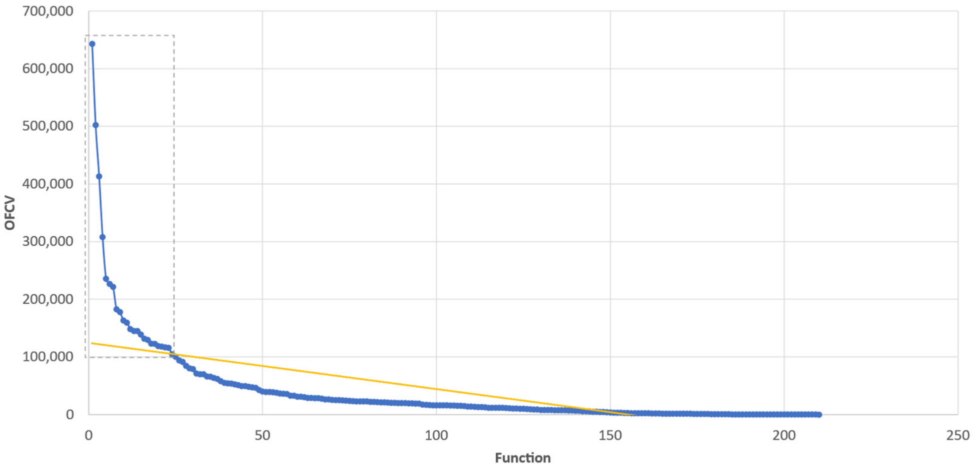


Figure 6. Scree test and regression curve using *OFCV* for the human driver. The first 23 functions are highlighted by a grey dashed box indicating risk functions.

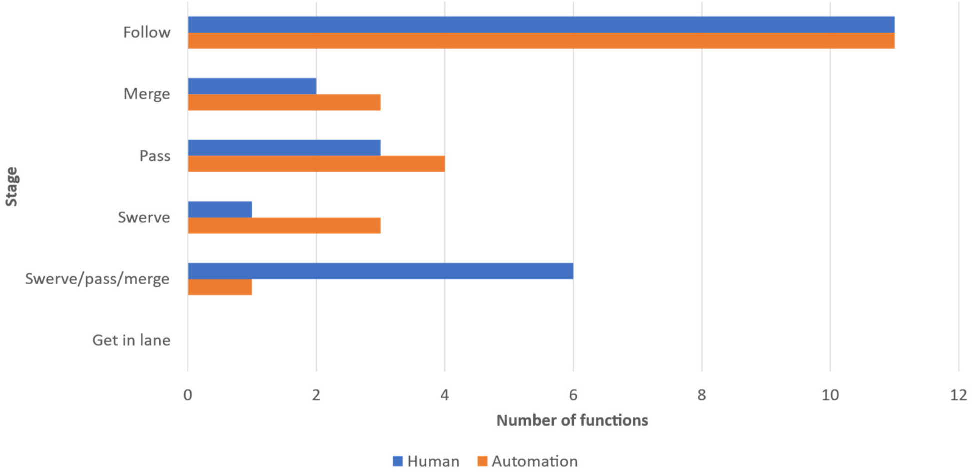


Figure 7. Comparison of the risk distribution over the stages between human and automation.

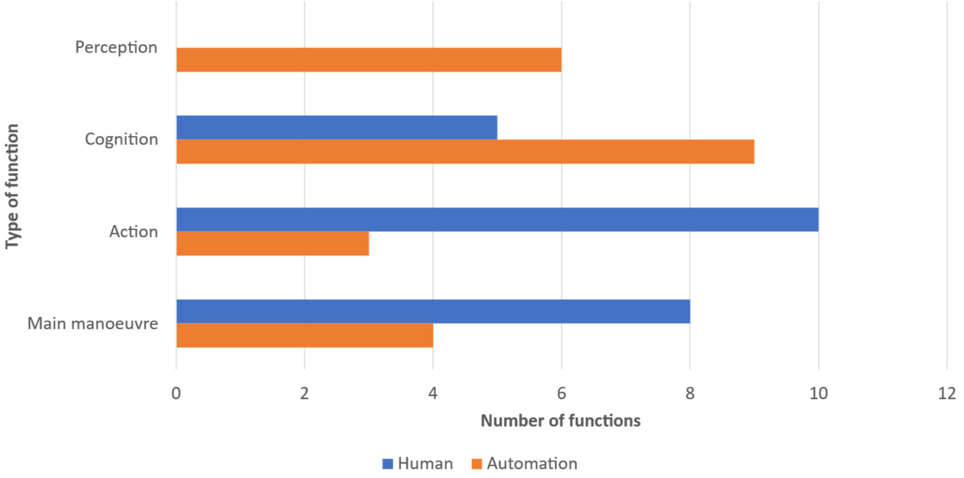


Figure 8. Comparison of the risk distribution over the types of function between humans and automation.

4.2.2. Analysis of Global System Variability

Finally, the *GSV* of each stage between human and automation is compared, as well as the function types for the EV in each stage. Figure 9 shows the comparison of *GSV* between humans and automation, where the variability is calculated in relation to the number of functions in the stage so that they can be compared relatively. The highest variability for both is found in the Pass stage and the largest difference between humans and automation occurs in the Follow stage, where the automation's variability is much larger than for humans. The other stages are relatively balanced, although the variability in automation is slightly lower. In general, automation has a higher overall variability.

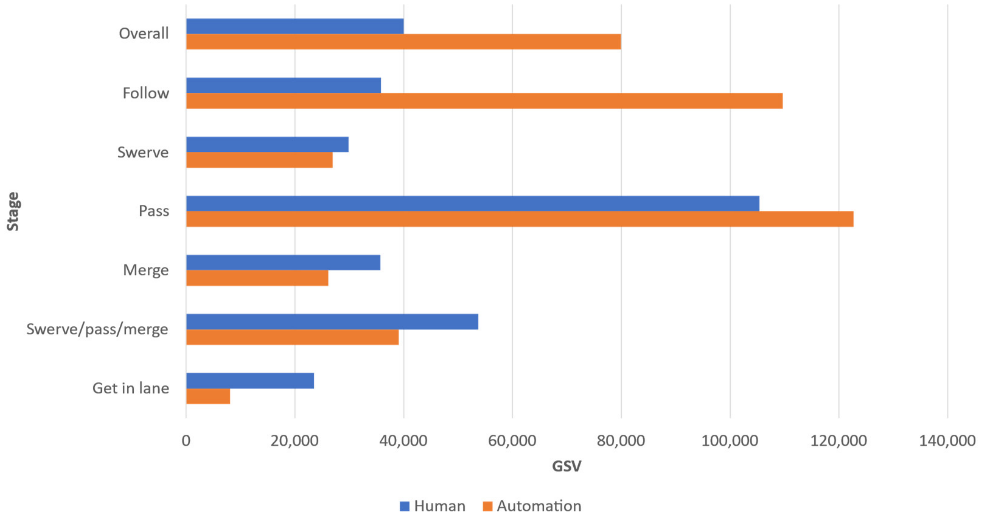


Figure 9. Comparison of the *GSV* for the overall system and per stage between human and automation.

4.2.3. Distinguishing the Interaction and Variability of System Functions for Potential Critical Functional Resonance

The previous analysis was very focused on the *OFCV* of risk functions and the *GSV*, which reflect the criticality of the functional variabilities in the system in an aggregated, abstract and simplified form. However, this criticality is composed of two dimensions: the variability a function receives (*ULFCV*) and transfers (*DLFCV*), which represent the functional variability, and the system resonance of a function, which reflects the interaction in the system, is how the functional variability is affected by the system (*WaD*) and how it influences the system itself (*WaU*). Therefore, these two dimensions were analysed separately for the system functions as well as risk functions in the following to get a deeper understanding. This is proposed by a matrix that represents the criticality of functions and their potential for functional resonance along the two dimensions functional variability and system resonance, which make up Functional Variability-System Resonance Matrix (FVSRM) (see Figure 10), a modification of the Variability Impact Matrix presented by Patriarca et al. [45]. For each function, the FVSRM considers in the system resonance dimension the sum of the *WaU* and *WaD*: low system resonance if it is lower than 5% of the maximum of the sum of *WaU* and *WaD*, medium system resonance if it is between 5%-30% of the maximum, and high system resonance if it is higher than 30% of the maximum. The functional variability dimension is considered by the sum of the *DLFCV* and *ULFCV*, where the three thresholds are analogous to the first dimension. The thresholds for both dimensions were determined subjectively by SMEs, inspired by the procedure of Patriarca et al. [45].

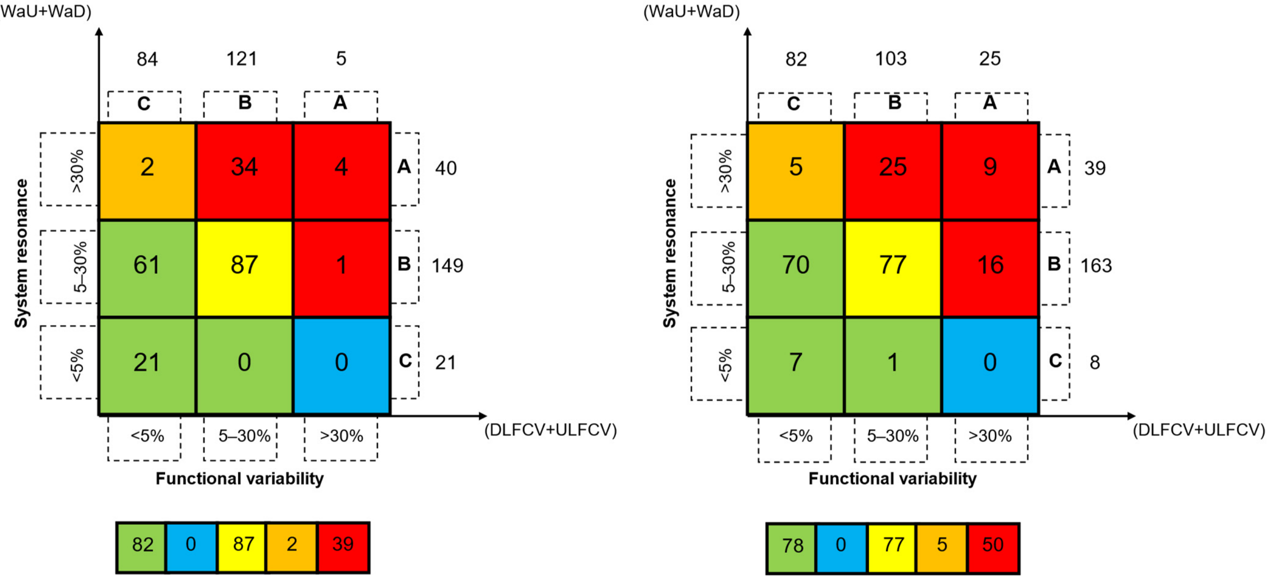


Figure 10. The Functional Variability-System Resonance Matrix (FVSRM), left for the human driver and right for the automation.

The FVSRM shows different areas: green (C-C, C-B, B-C) for uncritical functions, blue (A-C) for high variable functions with low system resonance, yellow (B-B) for medium variable functions with medium system resonance that are between uncritical and critical functions, orange (C-A) for low variable functions with high system resonance and red (B-A, A-A, A-B) for critical functions. Here, the orange and blue areas refer to functions that must be viewed with caution due to their special features. Functions in the blue area are functions that are typically error-prone but usually remain without adverse consequences (i.e., accidents) because they have a low systemic resonance. Functions in the orange area are functions where errors rarely occur, but when they happen, a strong systemic effect and consequently a high probability of accidents must be expected. In general, the functions in the orange area pose a greater hazard than the blue ones and are thus to be assessed as more critical. Below the FVSRM, the sum of functions per area is presented. Furthermore, the sum of functions per row and column is given to reflect the number of functions per dimension category.

The distribution of the functions in the FVSRM in Figure 10 shows that the system for the human driver is generally stable in terms of variability as five functions are above 30% functional variability but is affected by several interrelated functions with great system resonance impacts as 40 functions are above 30% system resonance. Instead, the distribution of the functions in the FVSRM for the automation is significantly more unstable in terms of variability as 25 functions have a functional variability of greater than 30%. Overall, the automation shows higher variable and medium system resonance functions. The number of uncritical functions is nearly the same for both at about 40%, with critical functions outweighing humans (19%) for automation (26%).

The risk functions for human drivers and automation were also analysed in a more differentiated way concerning the two dimensions of functional variability and system resonance, see Figures 11 and 12. Figure 11 shows the functional variability (*DLFCV* and *ULFCV* as stacked columns, left y-axis) and system resonance (*WaU* and *WaD* as stacked line markers, right y-axis) of risk functions (x-axis) for the human driver and Figure 12 for automation. Additionally, the thresholds for high functional variability and high system resonance are marked by the two dashed red lines. Some risk functions for the human driver are highlighted and explained below. The red highlighted functions are most critical because they have a high functional variability combined with high system resonance. Here, < maintain headway separation (EV) > and < follow LV (EV) >, in particular, stand out, with high variability and system resonance values, whereby they transfer variability for the most part and receive very little. In addition, each critical risk function is an action task. The orange highlighted functions are risk functions that have relatively low variability but combined with a strong system resonance. It can be argued that these functions are success factors demonstrating resilience because, despite their strong system effect and affectedness, they have little variability and are therefore stable. In particular, < driving free (OV) > and < driving free (LV) > with very high system resonances are noteworthy here. These functions must nevertheless be viewed with caution, especially under different scenario conditions, as a sudden increase in variability in these functions may have a large systemic effect. The function < assess opportunity to overtake safely (EV) > is also special because it is strongly influenced by the system and receives a relatively large amount of variability, but transfers very little variability into the system. Further, the functions < assess opportunity to overtake safely (EV) > and < merge back into starting lane (EV) > exhibit fairly high system resonances, but with relatively low variability. So, errors rarely occur here, but if they do, then they often result in accidents. Risk functions, either high variability combined with low system resonance or low variability joined with low system resonance, do not exist. By contrast, the latter is logical, otherwise, they would not be considered as risk functions.

Compared to the automation in Figure 12, it can be seen that humans have significantly lower variability values and that overall, significantly more risk functions in automation have high functional variability. However, the values of the system resonance are slightly higher for the human risk functions than for automation ones.

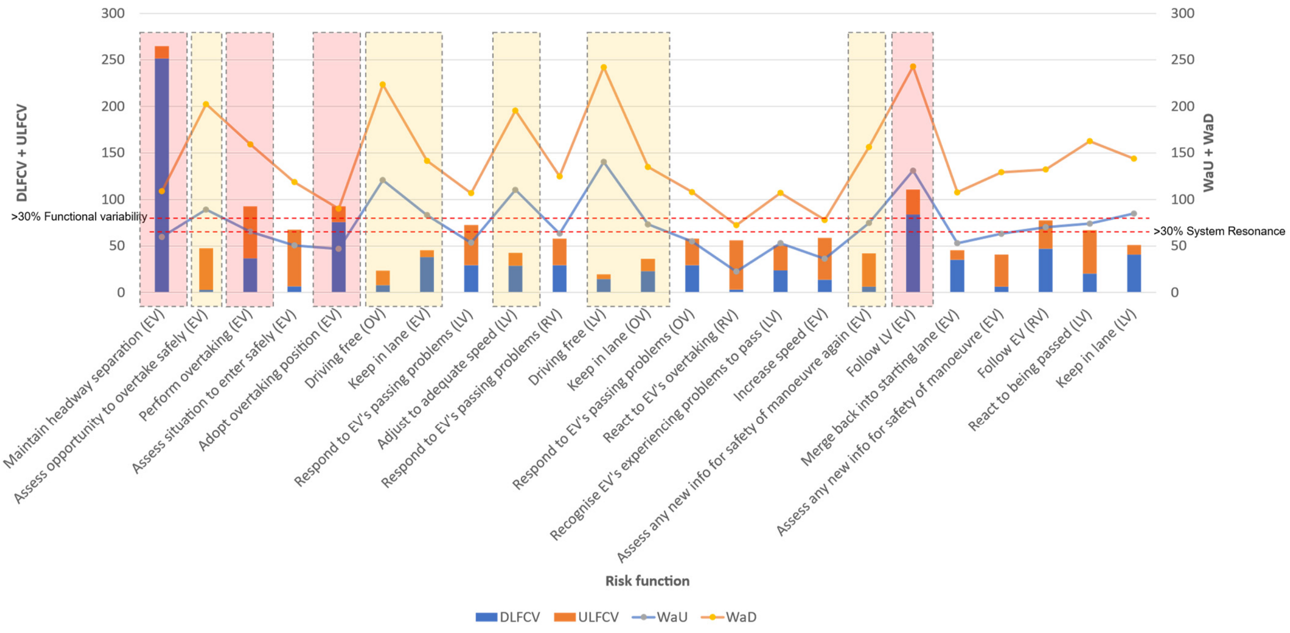


Figure 11. Risk functions for human drivers composed of functional variability and system resonance.

Several risk functions are also colour-coded in the automation (see Figure 12). This results in seven critical functions (red), with < observe oncoming traffic (EV) > standing out. Conspicuous compared to the human driver is the distribution of critical functions among the function types: five cognitive tasks, one perceptual task, and only two action tasks. Furthermore, four risk functions can be identified as success factors (orange), for example < follow LV (EV) > and < keep in lane (LV) >, each with high systemic resonance and low variability. In addition, there are risk functions in automation that have a relatively low systemic impact but are highly variable (blue), especially < watch for hazards located at roadside environment (EV) > or < assess road conditions (EV) >. It can be argued that these high functional variabilities are somewhat irrelevant because of their low system resonance, and therefore, they rarely lead to adverse events. Nevertheless, this variability should not be underestimated, especially if the scenario conditions change and thus the system resonance may change.

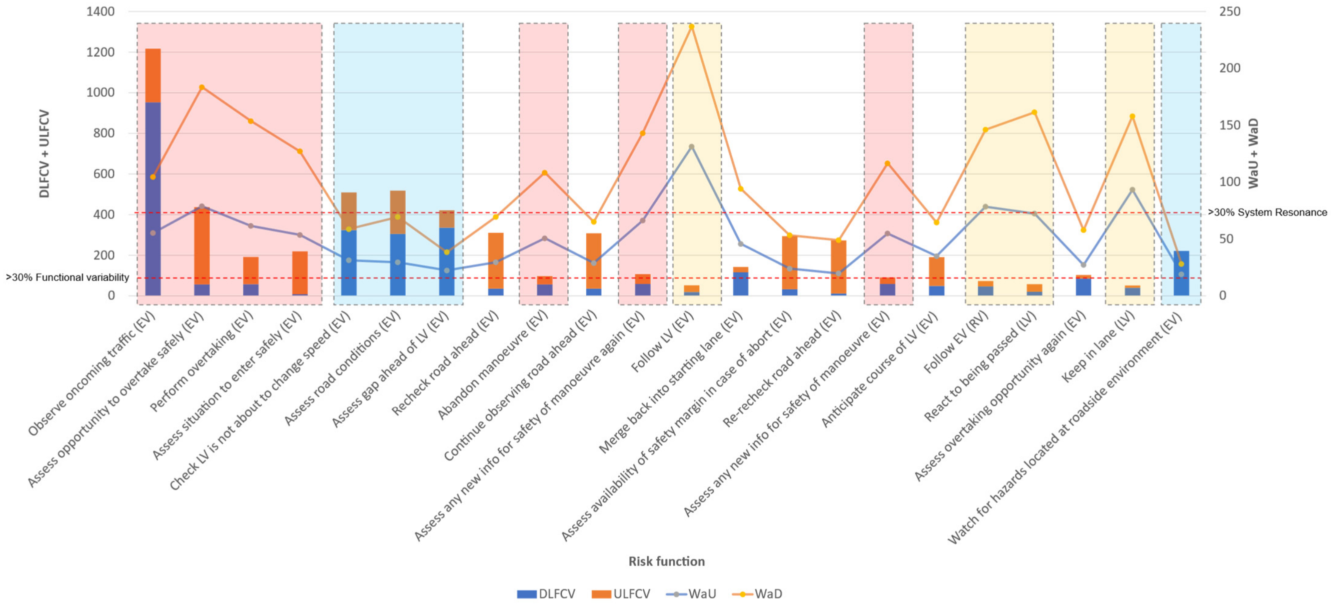


Figure 12. Risk functions for automation composed of functional variability and system resonance.

4.2.4. Analysis of Critical Paths

The quantitative evaluations shown previously were used to obtain an overview of the influence of system functions and their variabilities and interactions in the system in comparison between human driver and automation. Finally, this information was qualitatively reflected in the model to enable the mechanisms to be fully understood. In the following, this is exemplified by one critical path each for the human driver and the automation. In this work, a critical path is defined as the direct couplings between a risk function and its upstream and downstream functions, which is why all indirect couplings are hidden, except the couplings between the direct upstream and downstream functions.

Figure 13 shows the critical path of the function < maintain headway separation (EV) >, which is highlighted in light blue and will be referred to in the following as function in focus 1 (FiF1), for the human driver with respective agents and stages. The upstream couplings are highlighted in orange and the downstream couplings in blue. Additionally, every function's hexagon belonging to the orange or red area according to the FVSRM is marked with a sine curve indicating critical functions. Additionally, the types of functions are labelled by the respective colours, as mentioned in Section 4.1.

The FiF1 has five uplinks with little incoming variability and twelve downlinks transferring a high variability output, solely in the Follow stage. The uplinks come from four EV functions and one LV function, which are all action functions. Interestingly, four of the five upstream functions are critical, since they receive a relatively large amount of variability, which, however, is not transferred to very much. In addition, it is noticeable that < keep in lane (EV) > is temporally connected with FiF1 and thus two critical functions are executed simultaneously, inducing a potential higher workload. The downlinks go predominantly to RV (9), so RV is strongly influenced by FiF1. Otherwise, this offers great potential for resilient system behaviour, in that RV can dampen the received variability through adapted behaviour. Only one downlink goes to LV and two to EV itself, whereby a direct feedback loop between < increase speed (EV) > and FiF1 is created, so the two functions can mutually resonate. Moreover, the downlinks are predominantly associated with action functions (7) and few with perceptual (3) or cognitive functions (2). In general, the FiF1 has low *intrarelatedness* but high *interrelatedness* (3rd highest); in particular, the upstream function < keep in lane (EV) > and downstream function < follow LV (EV) > also have very high *interrelatedness*, so they form a "strongly interacting function triangle" here. It can be said that overall, the critical path of FiF1 is very action-heavy, has high interaction with other agents, a lot of variability accumulates in and around FiF1 (due to high *CTV*), and FiF1 has a strong system effect but is relatively little affected.

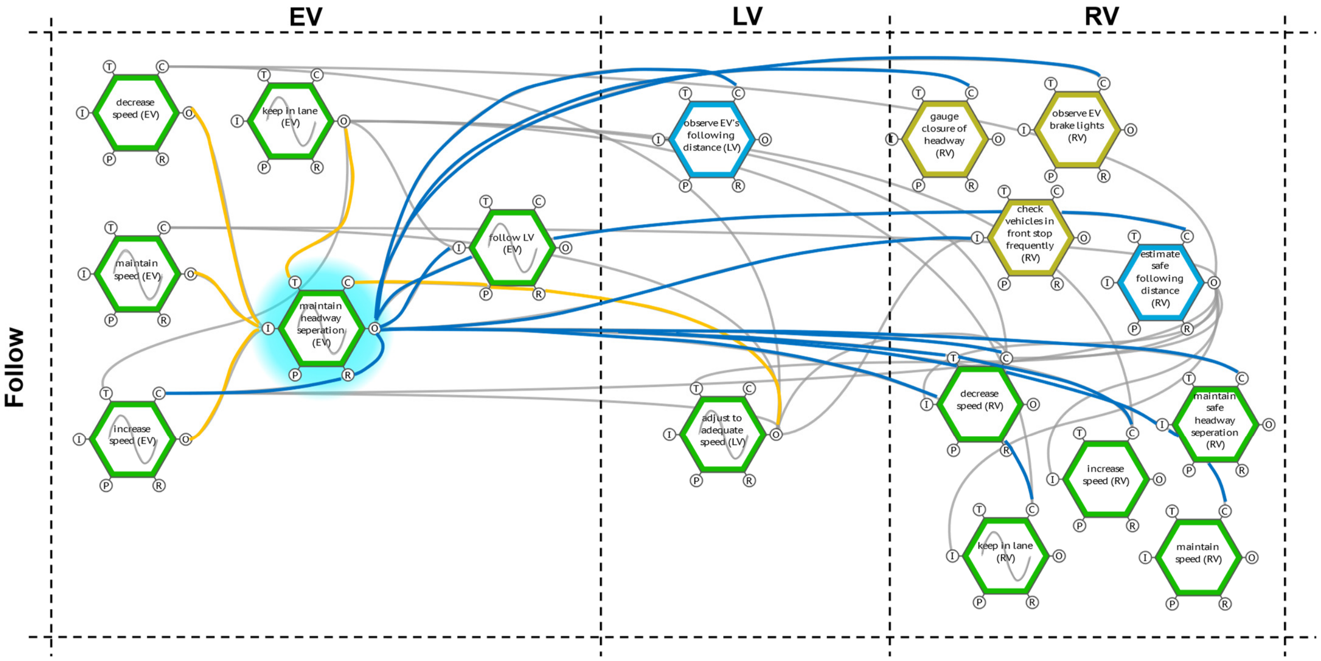

EV
LV
RV
Follow
decrease speed (EV)
keep in lane (EV)
observe EV's following distance (LV)
gauge closure of headway (RV)
observe EV brake lights (RV)
maintain speed (EV)
follow LV (EV)
check vehicles in front stop frequently (RV)
estimate safe following distance (RV)
maintain headway seperation (EV)
increase speed (EV)
adjust to adequate speed (LV)
decrease speed (RV)
maintain safe headway seperation (RV)
increase speed (RV)
keep in lane (RV)
maintain speed (RV)

Figure 13. The critical path of the function < maintain headway separation (EV) > for the human driver.

Figure 14 depicts the critical path of the function < observe oncoming traffic (EV) >, which is highlighted in light blue and will be referred to in the following as function in focus 2 (FiF2), for the automation with respective agents and stages. The FiF2 has six uplinks with high incoming variability and eleven downlinks transferring a high variability output, mostly in the Follow stage and less in the swerve and merge stages. The uplinks come from five EV functions and one OV function, which are four cognitive functions, one perception, and one action function. Interestingly, the distribution of upstream variability is very different with 60% coming from < determine pass can be completed (EV) > and < observe for lurkers behind OV (EV) > (30% each), and the rest coming from < judge available passing time (EV) > (18%), < judge LV's relative speed to OV (EV) > (11%), < judge distance from first OV (EV) > (10%), and < driving free (OV) > (1%). The downlinks go merely to EV's functions and predominantly to the Follow stage (7), only two downlinks go to each of the swerve and merge stages. In particular, the FiF2 is temporally coupled with five downstream functions, that is < assess road conditions (EV) >, < check LV is not about to change speed (EV) >, < assess gap ahead of LV (EV) >, < anticipate course of LV (EV) >, and < judge speed and performance of EV (EV) >, and thus six functions are executed simultaneously. In particular, most of these downstream functions also have a highly variable output and they are all received as an input in < assess opportunity to overtake safely (EV) >, which in total offers great potential for functional resonance. Moreover, the downlinks are predominantly associated with cognition functions (8) and few with perceptual functions (3). In general, the FiF2 is mainly connected to critical functions (except two functions) with high *intrarelatedness* but low *interrelatedness*. It can be said that overall, the critical path of FiF2 is very cognition- and perception-heavy, has high interaction within an agent over different stages, a lot of variability accumulates in and around FiF2 (due to high *CTV*), and FiF2 has a strong system effect and also high system affectedness, making it a highly critical function within EV's operations by automation.

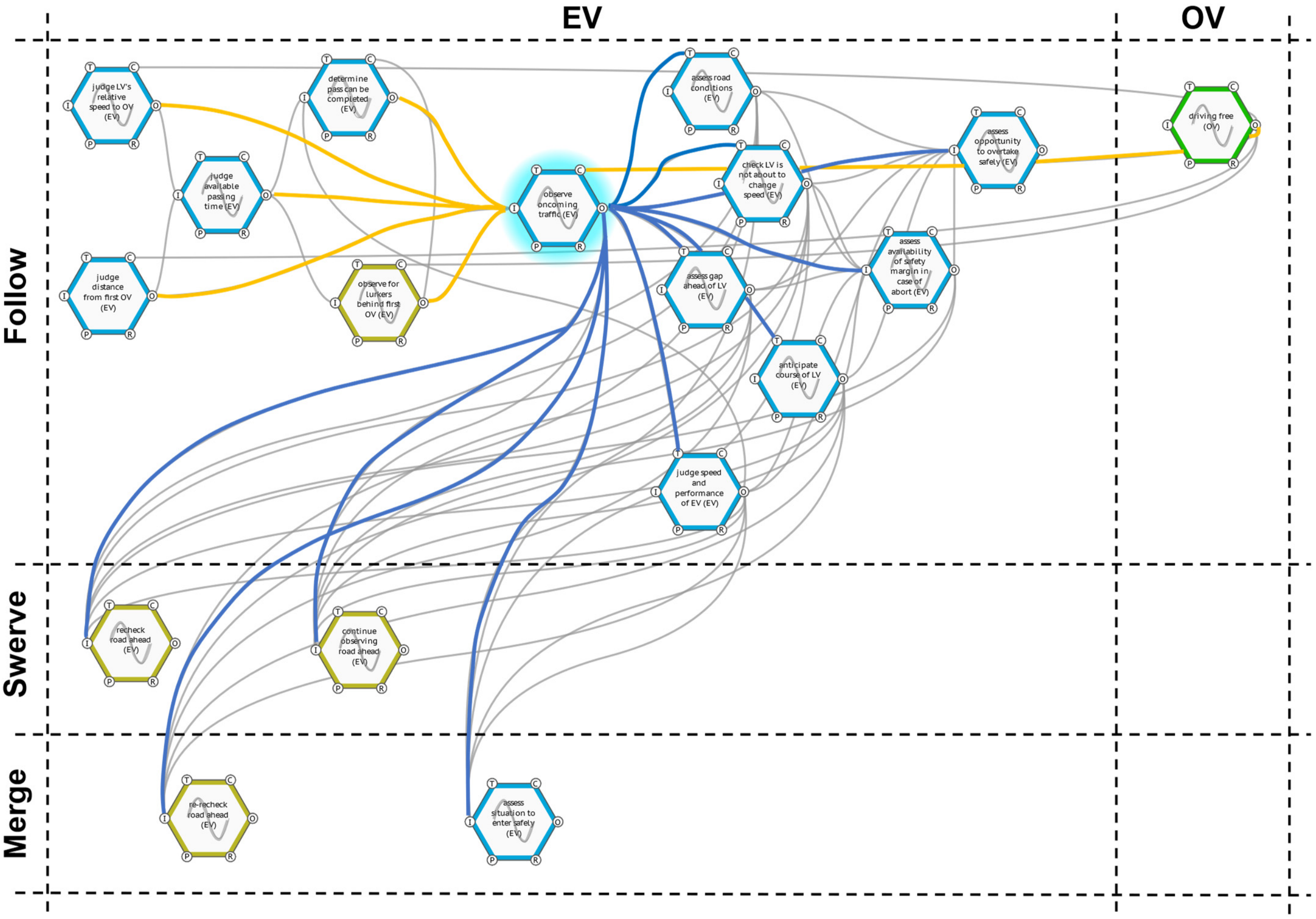

EV
OV
Follow
Swerve
Merge
judge LV's relative speed to OV (EV)
determine pass can be completed (EV)
judge available passing time (EV)
judge distance from first OV (EV)
observe for lurkers behind first OV (EV)
observe oncoming traffic (EV)
assess road conditions (EV)
check LV is not about to change speed (EV)
assess opportunity to overtake safely (EV)
driving free (OV)
assess gap ahead of LV (EV)
assess availability of safety margin in case of abort (EV)
anticipate course of LV (EV)
judge speed and performance of EV (EV)
recheck road ahead (EV)
continue observing road ahead (EV)
re-recheck road ahead (EV)
assess situation to enter safely (EV)
T
C
I
O
P
R

Figure 14. The critical path of the function < observe oncoming traffic (EV) > for the automation.

4.3. Recommendations for System Design and Validation

Based on the previous analyses, this subsection deals with recommendations for system design concerning the EV functions to improve the safety of the overall traffic system, as well as for validation focus of automation to reduce the test effort. First, a function allocation between human driver and automation is presented, followed by recommendations for automation's validation process.

4.3.1. Function Allocation between Human Driver and Automation

Automation of the entire scenario is not recommended, as automation is significantly more variable than humans in global system variability. However, the individual stages where automation is less variable could be automated in the sense of an authority transfer. The Follow and Pass stages would then be carried out by humans, and the Swerve, Merge and Get in lane stages by automation. With this approach, however, the individual functions are not considered and the automation of certain functions per stage would represent a more differentiated approach based on the compensatory design principle for automation according to Fitts [77], see Figure 15. Here, the function allocation for EV between humans and automation is shown. The driving tasks are divided according to stages and function types within the stages. The driving tasks are performed by the human (blue), by the automation (orange), or by both in the sense of shared control (grey), which is depicted both as a percentage and as an absolute value. In this paper, shared control means that the human and the automation work in collaboration simultaneously to achieve a single function [78] as an extension, that is, the capabilities of the human are extended by the automation or vice versa [79].

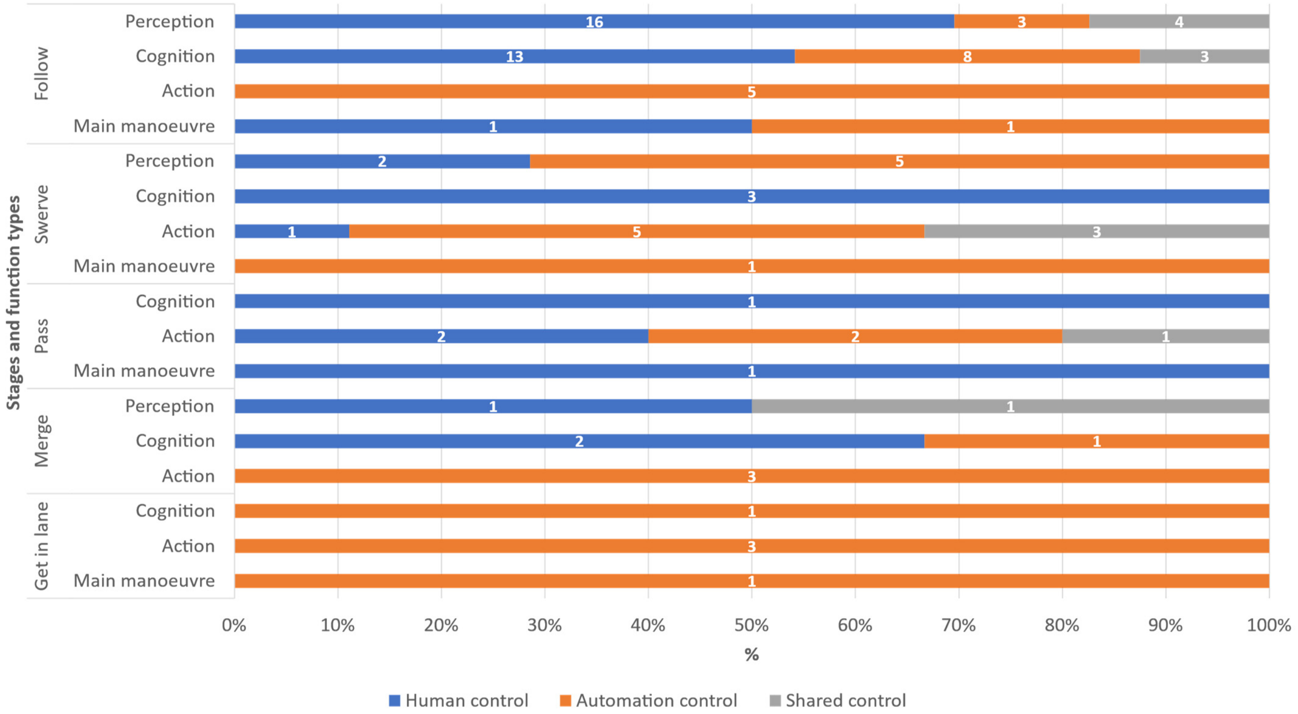


Figure 15. Function allocation between human driver and automation.

The decision about the assignment of the functions is based on the previous quantitative as well as qualitative analyses and the comparison of the functional variability and system resonance of each EV's function between humans and automation. If there was no clear and significant difference regarding the main performance indicators in a specific function between the human and automation, further metrics from Section 3.5, as well as the interaction with other functions and their performance indicators (see Section 4.2.4), were used.

First of all, it is noticeable that in the Follow and Pass stage, most of the functions are executed by humans and in the other three stages, the majority are executed by automation. The last stage in particular is performed exclusively by automation. Only 12% of all functions are executed as shared control, whereby this can take place at all three information processing levels. With the types of function, it is noticeable that humans perform significantly more perception and cognitive functions than automation, except in the Swerve or Get in lane stage, respectively. Action functions, on the other hand, are carried out significantly more by automation. Two of the five main manoeuvre functions should primarily be carried out by the human driver. These are the decision to overtake and the overtaking manoeuvre itself. The other three (following the lead vehicle, adopting the overtaking position, and completing the overtaking manoeuvre) are primarily related to automation.

The presented design recommendations for function allocation between human driver and automation can be seen as a joint cognitive system (JCS) [80] that regards human and machine as equal partners collaborating in the sense of a human-machine coagency "by shifting the focus from human and machine as two separate units to the JCS as a single unit" [80] (p. 67). This coagency is expressed in terms of function-centeredness [81] where system functions of the EV needed to accomplish the overtaking manoeuvre are distributed between the human driver and/or the automation in consideration of the interactions and dynamics in the system (reflected by system resonance) and the functional variabilities. In terms of SAE 3016, the resulting concept could also be realised as a highly assisted driving system instead of automated driving.

4.3.2. Validation Focus of AD

For the automation validation process of AD with the assumption of automating the whole scenario and its associated functions, particular attention should be paid to the risk functions for automation (see Appendix D in Table A4). This especially applies to functions in the Follow and Pass stages, as well as those that are declared perceptual and cognitive tasks. In addition, the validation focus can be expanded to include the critical functions in the red and orange areas of the FVSRM in Figure 10. The validation process can likely be reduced to the testing of these functions, such as criteria for exclusion, to reduce the test effort. This has to be fulfilled by AD. Otherwise, we do not even need to carry out further tests.

Otherwise, the function allocation shown in Figure 15 could be used to validate merely the functions in which automation is responsible alone or together with humans, and thus in turn reduce the validation effort to a level similar to current advanced driver assistance systems or SAE-Level 2 vehicles, where humans are completely responsible for the safety of the driving task. The only difference is that humans are not responsible for all functions, but only those allocated to them, and thus, automation takes responsibility for several other functions.

5. Discussion

This paper aims to identify and compare road traffic mechanisms in an overtaking scenario between a human driver and a highly automated vehicle, using FRAM. Based on this, the contributions of both agents regarding the safety of the overall system can be evaluated in order to derive system design recommendations for AD and insights to reduce the effort involved in the validation process. Thus, the results have to be interpreted and reflected upon, and the methodological application of FRAM must be discussed.

The results of the system design recommendations, including the function allocation between human driver and automation, suggest that complete automation of the overtaking scenario as a generic concept is currently unrealistic and inadvisable. Rather, humans must be more or less engaged in the driving task, especially for perception and cognition functions, until reliable full automation is implemented. This recognition is emphasised by Zhang et al. [82], who recommend not pursuing a narrow role for the human driver as a passenger or, at most, a fallback at an operational level according to the three control levels of driving by Michon [83], but rather holistically exploring other opportunities and roles for human drivers such, as a “commander role” at strategic and tactical levels, e.g., [84–87]. This is also in line with the design and effect space of shared control and human-machine cooperation conceptualised by Flemisch et al. [88], or the multi-level cooperation proposed by Pacaux-Lemoine and Flemisch [89]. Therefore, the short- and midterm strategy for automation in the overtaking scenario on rural roads to improve traffic safety should be to pursue a JCS approach for the traffic system [90] realising a human-automation collaboration and coagency throughout the driving scenario to achieve their common goal, which is to overtake safely. Thus, a differentiated approach must be taken that is centred on functions [81], whereby the functions of the JCS are divided according to different function types [76] and then functions are allocated to the agents, based on the FRAM analysis in Section 4, in the sense of “who does what”. This is in contrast to the six rigid levels of driving automation (LoDA) of the SAE and rather prefers as design decision of automation the view of the ten levels of automation (LoA) according to Sheridan [78] in combination with the four functional types by Parasuraman et al. [76]. This is also in line with the critique of the SAE’s LoDA definitions, especially conditional driving automation, by Inagaki and Sheridan [91]. In this paper, the function allocation between the two agents is a mix of shared control [78] and “static” trading of control [78], where static trading of control means that either the human or the automation is responsible for a function, and their role does not change from one occasion to another, or in different scenario conditions. Additionally, for reasons of simplicity, the extent of automation according to the LoAs is not considered. Unfortunately, this does not fit the real system behaviour perfectly, as technological changes lead to dynamics and adaptations in the functions by the human in collaboration with the automation. This can sometimes result in negative effects, such as the out-of-the-loop performance problem, loss of situational awareness, complacency or overtrust, or automation surprises, e.g., [92–96], so that there are eventually no positive changes as a net effect. A good example of this is the introduction of better brakes in the vehicle to increase road safety, assuming that the driver continues to drive as usual. However, his or her driving behaviour changes with the better brakes in that the driver drives faster because he/she can brake harder [80], which can be explained by the risk homeostasis of Wilde [97]. Maybe too strong an allocation or fragmentation of the functions makes little sense, since individual functions have to be carried out as a whole, sometimes well trained unit by one agent, otherwise too much information is missing or the information cannot be efficiently and effectively transferred at the interface between humans and automation. Thus, for the future, it would be more appropriate to implement an adaptive automation system [79] or a function-congruence [98] in the sense of "who does what and when", where functions can be shared or traded between humans and automation in response to changes in situations or human performance [79]. However, it must also be considered that drivers are usually not well trained, and such a complex function allocation could lead to confusion besides advantages. Therefore, in future research, the FRAM model for the overtaking scenario and the current design recommendations should be checked by "what-if analyses" [99,100] as various instantiations of the FRAM model in other scenarios (for example in curves or bad weather conditions) on the one hand, and on the other hand for dynamic performance changes over time, such as by Hirose et al. [57]. Furthermore, it is not only the performance variability that can change but also new functions will emerge through the collaboration between humans and automation, which is why an adaptation of the FRAM model in relation to the context conditions is necessary. For this purpose, in the future, the performance indicators per function must also be recalculated for the system with the new allocation of functions and iteratively adjusted because of the effect of contextual factors. Overall, the current design concept fits the basic scenario analysed well and is a good starting point but is not generally applicable and has to be adapted in further iterative analyses, both in theory and in practice.

Furthermore, the results as positive and negative contributions of the human driver and automation to system safety, as described in Section 4.2, need a comparison with the state-of-the-art knowledge regarding this issue. A thorough review would go beyond the scope of this paper, which is why only a comparison of the fundamental facts is described below. Unfortunately, the comparison will predominantly focus on the negative contributions of the human driver, as this is where large data have been analysed in the past. Whereas no substantial knowledge about the positive contributions of the human driver exists because data collections in the past and also currently focus on rare, critical, or even more rarely occurring accidents [13]. Therefore, the total number of successfully completed situations and the accidents currently successfully prevented by drivers is unknown, which is why ultimately information on uncritical situations cannot be found in the literature. This also coincides with the strong focus of the safety-I perspective in road traffic, as mentioned in the introduction. No comparison can be made for the automation either, as Level 4 vehicles have not been approved yet and only test drives are carried out in California. The data collected during the test drives have already been analysed, e.g., [101–103], but only on a relatively abstract level in the sense of defining causal reasons for disengagements or accidents such as system failures, road infrastructure, other road users, weather, etc., but not on a specific task level that would be required. Regarding the negative contributions of the human driver, the following can be found in the literature. According to Durth and Habermehl [104], most overtaking accidents occur in the Pass stage, with a proportion of 48%. This is in line with the calculated *GSV*, since the Pass stage has the highest variability and, therefore the greatest risk of accidents. According to Richter and Ruhl [62], the most common cause of overtaking accidents on rural roads in terms of fatalities is overtaking despite oncoming traffic, at 42%. The second most common cause of accidents is overtaking despite unclear traffic conditions (19%) and the third most common cause is errors when re-joining the right lane (14%). Interestingly, the first and third common causes of accidents can be identified by the critical functions < assess opportunity to overtake safely (EV) >, and < merge back into starting lane (EV) > which exhibit fairly high system resonances, but with relatively low variability. So, errors rarely occur here, but if they do, then they often result in accidents. The second most common cause of accidents could not be acknowledged by the results as < observe oncoming traffic (EV) > or < assess road conditions (EV) > do not pose a high risk for the human driver in the FRAM model. In addition, inappropriate speed, insufficient distances, and lack of attention are often contributing factors to accidents [13,105]. These factors can also be reflected by the critical functions < maintain headway separation (EV) > and < follow LV (EV) > which represent a mix of high speeds and low distances. However, the lack of attention cannot be confirmed because it is not explicitly stored as a function in the model and is rather implicitly included in other functions. These examples predominantly provide further evidence of the confirmability of the study by practising reflexivity, which in part increases the confidence in the validity of the FRAM model. If we set the former comparisons in relation to the results for the contributions of automation in this work, the following is noticeable. First, the high variability in the Pass stage also applies to the automation, even to a greater extent, which is why the automation does not provide support in this case. Second, the common accident causes of overtaking despite oncoming traffic or unclear traffic conditions, and errors when re-joining the right lane cannot be addressed by the automation either because of high variabilities in the functions < assess opportunity to overtake safely (EV) >, < observe oncoming traffic (EV) > or < assess road conditions (EV) >, and < merge back into starting lane (EV) >. Instead, the problem of inappropriate speeds and insufficient distances can be effectively tackled through automation, as the corresponding functions show low variability for the automation. As a result, it can be concluded that some known accident black spots are reflected in the results of the negative contributions by the human driver, many of which, however, cannot currently be improved by automation.

The results for the validation process of AD reveal insights for the potential reduction of test effort in two directions: First, assuming full automation, the identified risk functions for automation can be used as criteria for exclusion, or second, assuming a function allocation between human and automation, the validation process can be reduced to the allocated functions for automation. This change of perspective based on a safety-II and RE analysis opens up completely new possibilities for solving the approval trap [106]. This approval trap arose since current test methods are not economically or practically feasible for AD [107]. Here, research is being undertaken to create new test methods, paradoxically the safety assessment of common alternative approaches, e.g., [108,109] follows solely a safety-I perspective. This view, which is currently too one-sided, will probably lead to automation surprises, as already mentioned in the introduction. However, it is precisely here that this paper uses the safety II perspective with a holistic socio-technical approach to show solutions for identifying as many additional automation risks as possible in order to avoid this issue.

Ultimately, the methodological application of FRAM and potential limitations are discussed. The resulting FRAM model confirms both the large-scale complexity of the overtaking scenario and its interwoven interactions, as well as the inherent overwhelming complexity of the traditional FRAM. Here, the application of the Space-Time/Agency framework and the semi-quantitative approach supports the complex safety analysis and facilitates the identification of criticalities based on functional variability and their systemic interactions highlighting the contributions of human drivers and potential automation in order to derive system design recommendations for systemic corrective measures. Moreover, the FRAM model enhances the understanding of the systemic mechanisms by, for example, explicitly showing the space-time structure with which specific agent or agents interact and how they behave, as well as how this can ultimately result in positive and negative consequences.

The FRAM model is very profound, based on various sources and a calibration by peers, which makes a reliable behavioural model of the socio-technical system of the overtaking scenario for the intended analytical purposes. Nonetheless, the model does not claim to be complete, especially not for other analysis purposes, but it is a good basic model to use when further analysing, for instance, the influence of other environmental and scenario conditions or changes over time.

The peer workshop for the validation of the FRAM model generally works well, but lessons learned for future research include that the calibration process can be enhanced by the peers developing a FRAM model themselves and comparing it with the original one to achieve a deeper understanding. In addition, real accidents could be modelled as "Mini FRAMs" according to Bridges et al. [110], based on accident reports that also serve as a comparison about the logic of the overall model.

The variability was also determined based on two different sources to map reality as closely as possible. It should be noted regarding the human driver that the driving simulator study is well suited to assessing action functions at the operational level, such as lane-keeping or keeping safety distances, but that perception and cognitive functions are difficult to determine even with the support of eye-tracking. Structured interviews, as in Section 3.4, are more appropriate for this. Nevertheless, given the limited self-awareness of humans about their performance limits the usefulness of this approach. Further, the narrowed sample does not represent the entire driver population, which is why the comparison of performance variability between humans and automation in the paper is only valid to a limited extent. Whereby the sample size is generally sufficient for the narrower population, since, for example, a sample size of 20 test drivers is sufficient for testing the controllability of driver assistance systems according to ISO 26262 [111]. Concerning automation, too little data is currently available, which is why there are no alternatives to expert assessment. In the future, it could also be interesting to use cross-linked driving simulator studies to explicitly observe the interactions between multiple human drivers, automation, and/or joint human-automation and their resulting variabilities within one simulation.

The function identification process and the creation of the FRAM model, as well as the gathering of variability data, is very time- and resource-consuming. This raises some practical limitations for FRAM, which must definitely be improved in the future in order to overcome the current research-practice gap of systemic models and methods [112], especially FRAM. Here, on the one hand, researchers are currently applying systemic methods due to the current state-of-the-art and, on the other hand, many practitioners continue to apply sequential or epidemiological methods because of their ease of use or popularity despite known limitations. Frequently mentioned reasons for this are a difficult and time-consuming application [113], reduced model validation and usability, and a potential analyst bias [112]. One solution could be the IT framework for sharp-end operators' WAD data gathering through a mobile app proposed by Constantino et al. [114]. Overall, the practical applicability of FRAM, in general, has to be researched and improved, as claimed by [115]. Instead, the analysis of results runs relatively quickly due to matured software support.

The new metrics for the semi-quantitative approach introduced in Section 3.5 to better calculate and visualise each function's interactivity in the system, as well as its complex emergence effects in the system, served their purpose. However, their significance as an influencing parameter, especially concerning the composition of the weighting factors *WaU* and *WaD*, is currently a theoretical concept that has to be empirically validated in the future. Thus, their usefulness as a weight for system influence of functional variabilities to incorporate complex and dynamic behaviour is limited.

Moreover, the various aspects of the couplings were currently treated in the same way in the calculations, except for the propagation factor in Appendix A in Table A2. For the future, a more differentiated approach can be considered, showing potential different effects because of aspects not only qualitatively but also quantitatively.

6. Conclusions

This paper shows how FRAM can be used for a systemic function allocation between humans and automation considering the interactions and complex dynamics of functional variabilities in a space-time continuum within and between agents in the system based on an enhancement of quantitative outputs of FRAM. The analysis reveals that human drivers currently make a better overall contribution to the safety of the overall system in the simple overtaking scenario on a rural road than AD could. However, individual functions are emerging at each overtaking stage that offer great potential for increasing safety through automation, collaboration, or assistance. In particular, as long as no reliable full automation has been implemented, this means that the future automation strategy of the vehicle aiming to improve traffic safety should be more differentiated based on a JCS approach combined with function-centeredness aiming to incorporate the strengths of both the human driver and the automation according to adaptive automation of human-automation coagency. This contrasts with the current, inflexible approach to automate everything as much as feasible based on the six LoDAs by the SAE. In particular, this change in perspective may also simplify the validation problems of AD.

In the future, however, more research will have to be undertaken on how the results can be transferred to other driving scenarios and situations, how adaptive automation for overtaking can be explicitly implemented in practice, and what potential effects result from changes in scenario conditions or performance over time. Additionally, in this work, the traffic system in the overtaking situation and its performance are analysed from a single perspective, which is safety. However, AD should help to make driving not only safer but also more efficient and comfortable [116]. In addition, people as active passengers in the vehicle or passive interaction partners outside with the vehicle must be able to trust the automation and accept the new technology. Unfortunately, these different perspectives of the system performance are frequently viewed in isolation, also called siloed thinking, revealing only a part of what goes on [117]. However, these different views are mutually dependent, so in the future, their analysis will have to be synthesised according to Synesis [117], which involves the unification of different perspectives (safety, efficiency, and comfort, among others) into one analysis.

In conclusion, this paper confirms that RE, in particular FRAM, can be applied to the road traffic system to design automated driving functions proactively and holistically, or rather the joint driver-vehicle system, demonstrating the potential for supporting decision-makers to enhance safety enriched by the identification of non-linear, complex, and emergent risks rather than the linear cause–effect-related risks that are frequently the sole focus of safety and risk assessments at present.

Supplementary Materials: The following is available online at https://www.mdpi.com/article/10.3390/safety8010003/s1, FRAM Model Visualiser Data S1: Overall FRAM model.

Author Contributions: Conceptualization, N.G., A.G. and M.H.; methodology, N.G., A.G and M.H.; software, N.G., A.G.; validation, N.G., A.G. and M.H.; formal analysis, N.G., A.G. and M.H.; investigation, N.G., A.G. and M.H.; resources, N.G.; data curation, N.G., A.G. and M.H.; writing—original draft preparation, N.G.; writing—review and editing, N.G.; visualization, N.G.; supervision, K.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved post hoc by the Institutional Review Board of the Ethics Committee of the Technical University of Munich confirming that the study does not require a formal vote (protocol code 513/21 S and date of approval 23 August 2021).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: In this section, the authors are grateful for all subjects who participated in this study. In particular, we thank the peers for their dedicated work at the validation workshop. A special thanks go to our colleagues, Michael Rettenmaier and Lorenz Steckhan, for their fruitful discussions about the ideas presented in this paper. Further, we would like to thank the editing service KERN for proofreading the paper.

Conflicts of Interest: The authors have no relevant financial or non-financial interests to disclose. Additionally, no funds, grants, or other support were received.

Appendix A

In the following, the formulas for the remaining metrics from Section 3.5.1 are provided:

In the first step, a numerical score was assigned to each performance variability characteristic (see Table A1). The higher the score, the more variable the output. The variability of the upstream output j, OV_j was the product of these two scores (A1):

$$OV_j = V_j^T \cdot V_j^P \tag{A1}$$

where:

V_j^T represents the upstream output j score in terms of timing

V_j^P represents the upstream output j score in terms of precision

Table A1. Assignment of numerical values to the linguistic description of variability manifestation of the phenotypes timing and precision.

Variability Phenotype	Variability Manifestation	V_j^Tor V_j^P
Timing	Too early	2
	On time	1
	Too late	4
	Not at all	5
Precision	Imprecise	5
	Acceptable	3
	Precise	1

However, the upstream outputs V_j^T and V_j^P must be calculated as a frequency distribution since they were collected as a distribution in the study. The reason for this is that a static behaviour of a system function does not adequately reflect a real case, and thus should rather be dynamic. Therefore, P_{TE}, P_{OT}, P_{TL}, P_{NAA}, P_{PR}, P_A and P_I represent the percentage distribution of subjects of the variability values too early (TE), on time (OT), too late (TL), not at all (NAA), precise (PR), acceptable (A), and imprecise (I), respectively. The percentage values are between 0 and 1. These are then weighted by the numerical variability values from Table A1. The calculation was thus as follows (A2) and (A3):

$$V_j^T = P_{TE} \cdot V_j^T(\ TE) + P_{OT} \cdot V_j^T(OT) + P_{TL} \cdot V_j^T(TL) + P_{NAA} \cdot\ V_j^T(NAA) \tag{A2}$$

$$V_j^P = P_{PR} \cdot V_j^P(PR) + P_A \cdot V_j^P(A) + P_I \cdot V_j^P(I) \tag{A3}$$

Once assigned the variability score for the upstream output, the coupling variability (CV) of the upstream output j and the downstream function i (A4) as well as associated variability propagation factors a_{ij}^T and a_{ij}^P had to be specified (A5):

$$CV_{ij} = OV_j \cdot a_{ij}^T \cdot a_{ij}^P \tag{A4}$$

where:

a_{ij}^T represents the propagation factor for the upstream output j and the downstream function i in terms of timing

a_{ij}^{P} represents the propagation factor for the upstream output j and the downstream function i in terms of precision

Note that a_{ij}^{T} or a_{ij}^{P} may assume the following values:

$$\begin{array}{cc} 2 & \text{if the upstream output has an amplifying effect on the downstream function} \\ 1 & \text{if the upstream output does not affect the downstream function} \\ 0.5 & \text{if the upstream output has a damping effect on the downstream function} \end{array} \tag{A5}$$

The specification of the propagation factor was based on Table A2. As before, for upstream output, percentage distributions were also considered for propagation factors a_{ij}^{T} and a_{ij}^{P}. The calculation was thus as follows (A6) and (A7):

$$a_{ij}^{T} = P_{TE} * a_{ij}^{T}(TE) + P_{OT} * a_{ij}^{T}(OT) + P_{TL} * a_{ij}^{T}(TL) + P_{NAA} * \ a_{ij}^{T}(NAA) \tag{A6}$$

$$a_{ij}^{P} = P_{PR} * a_{ij}^{P}(PR) + P_{A} * a_{ij}^{P}(A) + P_{I} * a_{ij}^{P}(I) \tag{A7}$$

Table A2. Upstream/downstream propagation of variability, according to Patriarca et al. [118].

Upstream Output Variability		**Input**	**Precondition**	**Resource**	**Control**	**Time**
Timing variability of output	Too early	A/NE	A	NE/D	A	A
	On time	D	D	D	D	D
	Too late	A	A	A	A	A
	Not at all	A	A	A	A	A
Precision variability of output	Imprecise	A	A	A	A	A
	Acceptable	NE	NE	NE	NE	NE
	Precise	D	D	D	D	D

A = Amplifying, NE = No Effect, D = Damping.

Appendix B

In the following, the formulas for the remaining metrics from Section 3.5.2 are provided:

The number of downlinks of an upstream function j (N_{DL}^{j}) and the number of uplinks of a downstream function i (N_{UL}^{i}) specifies the number of links of an upstream function to downstream functions or vice versa. N_{DL}^{j} is the sum of downlinks of an upstream function j (A8) and N_{UL}^{i} is the sum of uplinks of a downstream function i (A9):

$$N_{DL}^{j} = \sum\nolimits_{i=1}^{j} DL_{ij} \tag{A8}$$

$$N_{UL}^{i} = \sum\nolimits_{j=1}^{i} DL_{ji} \tag{A9}$$

It should be mentioned that only the downlinks or uplinks between two foreground functions and not between two background functions or between a foreground and a background function were counted, as background functions are stable and not variable and represent the system boundary, which are therefore not included in the analysis.

Intra-stage links calculates the number of downlinks and uplinks of a function f where the linked upstream j and downstream functions i are in the same stage St and executed by the same agent Ag (A10):

$$\begin{gathered} Intra - stage\ links_{f} = [\sum\nolimits_{i=1}^{f}\ if\ ((Ag_{f} = Ag_{i}\ \&\&\ St_{f} = St_{i})\ then\ 1,\ else\ 0) + \\ \sum\nolimits_{j=1}^{f}\ if\ ((Ag_{f} = Ag_{j}\ \&\&\ St_{f} = St_{j})\ then\ 1,\ else\ 0)] \end{gathered} \tag{A10}$$

Intra-agent links calculates the number of downlinks and uplinks of a function *f* where the linked upstream *j* and downstream functions *i* are in different stages *St* but executed by the same agent *Ag* (A11):

$$\begin{gathered} Intra-agent\ links_f = [\textstyle\sum_{i=1}^{f} if\ ((Ag_f = Ag_i\ \&\&\ St_f \neq\ St_i)\ then\ 1,\ else\ 0)+ \\ \textstyle\sum_{j=1}^{f} if\ ((Ag_f = Ag_j\ \&\&\ St_f \neq\ St_j)\ then\ 1,\ else\ 0)] \end{gathered} \tag{A11}$$

Intrarelatedness calculates the interaction within an agent and results from the sum of the *intra-stage links* and the *intra-agent links* of a function *f* (A12):

$$Intrarelatedness_f = Intra - stage\ links_f + ss{\cdot}Intra - agent\ links_f \tag{A12}$$

where the *intra-agent links* were additionally weighted by a factor β, since a link of a function to another stage has a higher system effect, and thus must be weighted more heavily. The chosen value for β in this work is 2.

Inter-agent links calculates the number of downlinks and uplinks of a function *f* where the linked upstream *j* and downstream functions *i* are executed by different agents *Ag* (A13):

$$Inter - agent\ links_f = [\sum_{i=1}^{f} if\ ((Ag_f \neq Ag_i)\ then\ 1,\ else\ 0) + \sum_{j=1}^{f} if\ ((Ag_f \neq Ag_j)\ then\ 1,\ else\ 0)] \tag{A13}$$

Moreover, the sum of the *intra-stage links*, *intra-agent links,* and *inter-agent links* is equal to the sum of the N_{DL}^{j} and N_{UL}^{i} for each function.

Different-linked agents calculates with how many different agents *k* a function *f* is directly connected through its upstream *j* and downstream functions *i* (A14):

$$\begin{gathered} Different\ linked\ agents_f = [\textstyle\sum_{k=1}^{4} if(\textstyle\sum_{i=1}^{f} if\ ((Ag_f \neq Ag_i\ \&\&\ Ag_i = Ag_k)\ then\ 1,\ else\ 0))+ \\ (\textstyle\sum_{j=1}^{f} if\ ((Ag_f \neq Ag_j\ \&\&\ Ag_j = Ag_k)\ then\ 1,\ else\ 0))] \end{gathered} \tag{A14}$$

The *interrelatedness* of a function *f* calculates the interaction between agents and is the result of the product of *inter-agent links* and *different linked agents* (A15):

$$Interrelatedness_f = Inter - agent\ links_f{\cdot}Different\ linked\ agents_f \tag{A15}$$

Direct feedback loops mean that a downstream function *i* of a function *f* is also an upstream function *j* of the function *f* and vice versa. This results in a loop between these two functions, in which only two functions are involved. The calculation is as follows (A16):

$$\begin{gathered} Direct\ feedback\ loops_f = [\textstyle\sum_{i=1}^{f} if(Coupling(f,i)\ \&\&\ Coupling(i,f))\ then\ 1,\ else\ 0)+ \\ \textstyle\sum_{j=1}^{f} if(Coupling(f,j)\ \&\&\ Coupling(j,f))\ then\ 1,\ else\ 0)] \end{gathered} \tag{A16}$$

where *Coupling* is a function that gives as result 1 if there is a direct connection between function *f* and its upstream function *j* or its downstream function *i*.

Indirect feedback loops involve more than two functions. For example, function A calls function B, which in turn is connected to function C, which again calls function A, closing the loop. The function (*Loops* calculates all cycles in the model that contain the function *f* and are not direct (feedback loops of function *f* (A17):

$$Indirect\ feedback\ loops_f = \sum Loops_f \tag{A17}$$

Mean feedback loops indicates how many functions occur in the mean of the indirect feedback loops from the function *f* and is calculated as follows (A18):

$$Mean\ feedback\ loops_f = \frac{\sum\ Length\ (\mathrm{Loops}f)}{Indirect\ feedback\ loopsf} \tag{A18}$$

where *Length* calculates the number of functions per cycle, that is, the length of the cycle of function *f*.

The last three metrics mentioned are then integrated into the *feedback loop factor* (A19):

$$Feedback\ loop\ factor_f = Direct\ feedback\ loops_f + Indirect\ feedback\ loops_f \cdot Mean\ feedback\ loops_f \tag{A19}$$

The *CTV* was used to calculate how much variability accumulates around a function *f*. To do this, the *ULFCV* of the coupled upstream functions *j*, the *DLFCV* of the coupled downstream functions *i*, and the *DLFCV* and *ULFCV* of the function *f* were added together (A20):

$$CTV_f = DLFCV_f + ULFCV_f + \sum_{i=1}^{f} DLFCV_i + \sum_{j=1}^{f} ULFCV_j \tag{A20}$$

The *Katz-centrality* calculates the relative influence of a function. According to Falegnami et al. [71], this metric is the most suitable metric for function prioritisation in a FRAM model analysis. For all connections that are reachable both upstream and downstream by the function *f*, the *CVs* of the upstream function of the respective connections are added together. To mitigate the indirect influence of the functions, that is the farther away a function is located, the lower its influence, the distances to the individual couplings are considered and weighted with a factor α. It should be noted here that a direct connection has zero distance. The d_{iif} gives the distance of a downstream connection to function *f*, where *ii* denotes direct and indirect downstream functions. The d_{jjf} reflects the distance of an upstream connection to function *f*, where *jj* denotes direct and indirect upstream functions. The weight factor α and *Katz-centrality* are calculated as follows (A21)–(A23):

$$\alpha_{iif} = \frac{1}{d_{iif} + 1} \tag{A21}$$

$$\alpha_{jjf} = \frac{1}{d_{jjf} + 1} \tag{A22}$$

$$Katz - centrality_f = \sum_{ii=1}^{f} CV_{ij} * \alpha_{iif} + \sum_{jj=1}^{f} CV_{ij} * \alpha_{jjf} \tag{A23}$$

Incloseness- and *Outcloseness-centrality* indicate how centrally a node (i.e., a function) is located within a network. They each form the sum of the reciprocal distances to reachable functions, weighted by the *CV* of the respective upstream functions. *Incloseness-Centrality* only considers upstream functions *j* of function *f*. The number of upstream functions reachable from function *f* are represented by *n*. *Incloseness-Centrality* was calculated as follows (A24):

$$Incloseness - centrality_f = \frac{n-1}{\sum_{jj=1}^{f} (CV_{ij} \times d_{jjf})} \tag{A24}$$

In contrast, the *Outcloseness-centrality* only considers downstream functions *i* of function *f* and is calculated as follows (A25):

$$Outcloseness - centrality_f = \frac{n-1}{\sum_{ii=1}^{f} (CV_{ij} \times d_{iif})} \tag{A25}$$

Betweenness-centrality shows how often a function *f* occurs as the shortest distance between two other functions in the model (A26):

$$Betweenness - centrality_f = \sum_{ii \neq jj \neq f \in V} \frac{\sigma_{iijj}(f)}{\sigma_{iijj}} \tag{A26}$$

where σ_{iijj} and $\sigma_{iijj}(f)$ represent the number of the shortest distances between a function i and j and the number of the shortest distances between a function i and j, in which function f occurs, respectively. The V indicates the quantity of all functions in the model, and ii and jj define that indirect downstream and upstream functions were also considered.

The metrics N_{DL}^{j}, N_{UL}^{i}, Intrarelatedness, Interrelatedness, Feedback loop factor, CTV, Katz-, Incloseness-, Outcloseness- and Betweenness-centrality were then transformed into relative metrics ($Met^{relative}$), which reflect the effect of a function compared to all other functions within a metric in percentage. This ensures that all metrics can be used as an equal weight in further calculations. Here, $\mathrm{Met_f}$, a specific value of one metric of a function f, is divided by the sum of all values of one metric for all functions k. However, this would lead to values below 1. This is problematic because, with further calculations, comprising multiplications, the amount would decrease. For this reason, the percentage values are divided by the inverse of all functions N in the model in order to always ensure a value above 1. This ensures that the values are magnified in further calculations and the influence of a function thus becomes apparent. The calculation for $Met^{relative}$ was the following (A27):

$$Met_f^{relative} = \frac{\frac{Met_f}{\sum_{k=1}^{N} Met_k}}{\frac{1}{N}} \tag{A27}$$

Finally, these relative metrics were integrated into the Weight as Upstream (WaU) and Weight as Downstream (WaD) as shown in Section 3.5.2.

Appendix C

Table A3. A rough description of the main functions of the overall FRAM model per each agent and stage.

Stage	EV	LV	RV	OV
Follow	to follow LV through recognising the following situation, keeping the lane, and maintaining headway separation; to decide to overtake or not, which is mainly based on assessing the opportunity to overtake safely, judging whether overtaking is permitted, and evaluating the reasonableness for overtaking	to drive free by keeping the lane and adjusting adequate speed; to react to being followed by EV through observing EV's intention to overtake as well as its following distance	to follow EV through recognising the following situation, keeping the lane, and maintaining headway separation	to drive free by keeping the lane and adjusting adequate speed
Swerve	to adopt the overtaking position by lane keeping, reducing headway from the normal following, and adjusting the speed to that of LV; to swerve completely to the oncoming lane afterwards checking any hazards behind or in front, assessing the overtaking opportunity is still safe and using the left indicator	to detect EV's swerving into the oncoming lane; to maintain speed; to react to being passed by responding to potential passing problems of EV (optional)	to detect EV's swerving into the oncoming lane; to react to being passed by responding to potential passing problems of EV (optional)	to detect EV's swerving into the oncoming lane; to maintain speed; to react to being passed by responding to potential passing problems of EV (optional)
Pass	to perform the overtaking through accelerating LV decisively or merging back into starting lane if the manoeuvre is unsafe and abandoning the manoeuvre	to detect the passing vehicle in peripheral vision; to react to being passed by responding to potential passing problems of EV (optional)	to react to being passed by responding to potential passing problems of EV (optional)	to react to being passed by responding to potential passing problems of EV (optional)
Merge	to merge progressively into the starting lane by adjusting EV's speed in relation to other traffic, assessing the situation to enter safely, and using the right indicator	to prepare to provide a larger opening for EV to merge back; to react to being passed by responding to potential passing problems of EV (optional)	to prepare to provide larger space to LV in case of EV's manoeuvre abandoning or to catch up to LV; to react to being passed by responding to potential passing problems of EV (optional)	to prepare for braking; to react to being passed by responding to potential passing problems of EV (optional)
Get in lane	to complete the overtaking through positioning into the starting lane evaluating the driving situation, and resuming at the desired speed	to follow EV; to react to being followed by RV	to follow LV	to drive free

The wording "(optional)" means that this function or task is not necessarily fixed to the assigned stage and rather can be executed in the Swerve, Pass, or Merge stage or not at all if not required.

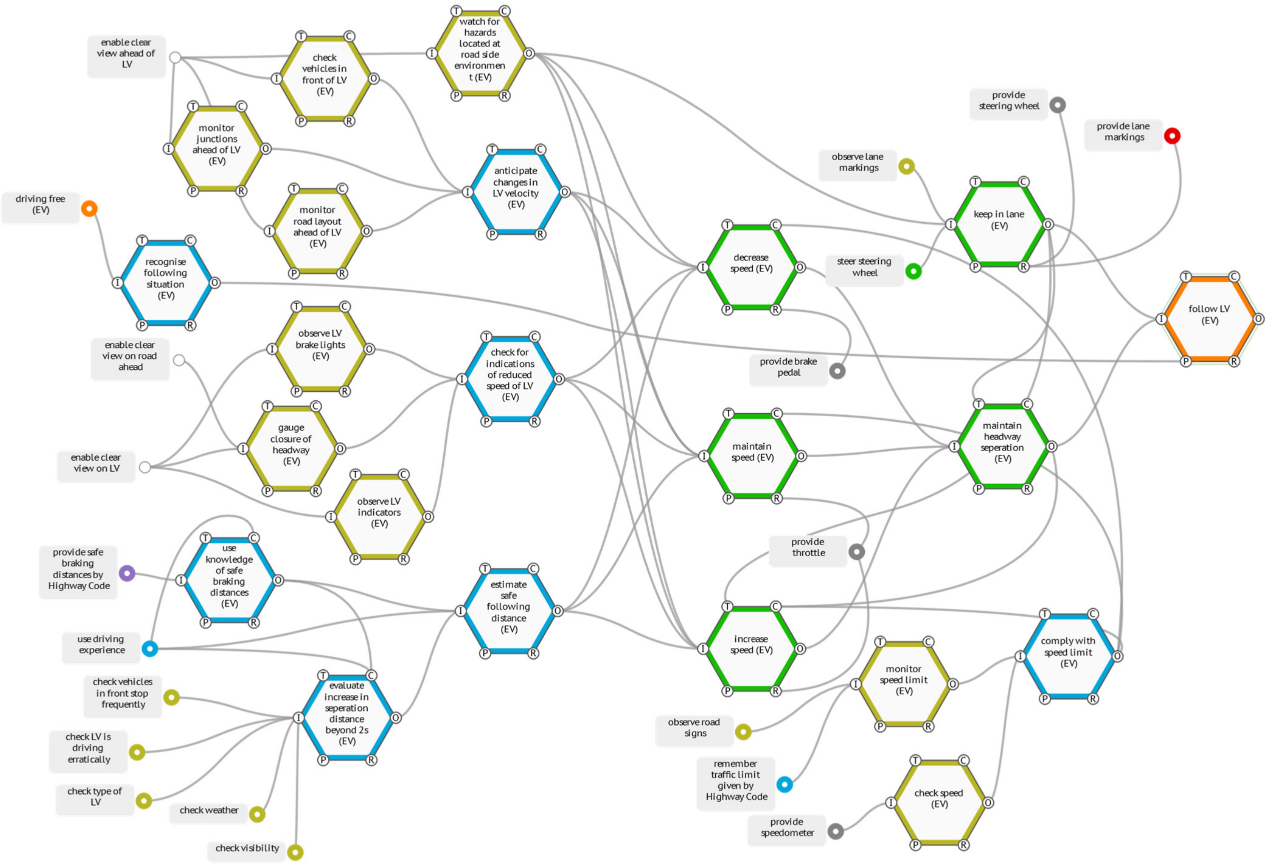

enable clear view ahead of LV
check vehicles in front of LV (EV)
watch for hazards located at road side environmen t (EV)
monitor junctions ahead of LV (EV)
anticipate changes in LV velocity (EV)
driving free (EV)
monitor road layout ahead of LV (EV)
recognise following situation (EV)
enable clear view on road ahead
observe LV brake lights (EV)
check for indications of reduced speed of LV (EV)
gauge closure of headway (EV)
enable clear view on LV
observe LV indicators (EV)
provide safe braking distances by Highway Code
use knowledge of safe braking distances (EV)
estimate safe following distance (EV)
use driving experience
check vehicles in front stop frequently
evaluate increase in seperation distance beyond 2s (EV)
check LV is driving erratically
check type of LV
check weather
check visibility
provide steering wheel
provide lane markings
observe lane markings
keep in lane (EV)
decrease speed (EV)
steer steering wheel
follow LV (EV)
provide brake pedal
maintain speed (EV)
maintain headway seperation (EV)
provide throttle
increase speed (EV)
comply with speed limit (EV)
monitor speed limit (EV)
observe road signs
remember traffic limit given by Highway Code
check speed (EV)
provide speedometer
T
C
I
O
P
R

Figure A1. Illustration of the following process of EV in the Follow stage of the overall FRAM model.

In Figure A1, the driving behaviour of to follow by EV in the Follow stage is explained in detail. Only foreground functions, as well as the couplings between the functions within EV and within the Follow stage, are explained and not connections to functions in other stages or agents. The explanation follows a reading of Figure A1 from right to left. The EV has to follow LV through recognising the following situation and keeping the lane and maintaining headway separation simultaneously. The headway separation is ensured by decreasing, maintaining, or increasing the speed, which are also regulated in compliance with the speed limit and headway separation. The driver complies with the speed limit by monitoring the speed limit as well as checking the speedometer. The speed regulation is further influenced by watching for hazards located at the road side, anticipating changes in LV velocity (based on monitoring traffic rules, road layout ahead and junctions ahead, and checking for vehicles in front of LV), checking indications of the reduced speed of LV (based on observing LV's brake lights and indicators as well as gauging the closure of headway) and estimating a safe following distance (based on using knowledge of safe braking distances and evaluating a required increase in separation distance beyond 2 s that is enabled by checking vehicles in front stopping frequently or whether LV is driving erratically). Furthermore, some functions are coupled with other agents or stages (not depicted in Figure A1). For example, keeping the lane or maintaining headway separation are influenced by the longitudinal and lateral driving behaviour of LV, and following LV is affected by LV's driving free performance or can also be influenced in the way if the assessment to overtake safely was judged as unsafe, then the following performance can be worsened through impatience.

Appendix D

Table A4. Risk functions for human driver and automation.

Risk Function	Human	Automation
Follow LV (EV)	x	x
Maintain headway separation (EV)	x	
Perform overtaking (EV)	x	x
Assess opportunity to overtake safely (EV)	x	x
Check LV is not about to change speed (EV)		x
Follow EV (RV)	x	x
React to being passed (LV)	x	x
Assess road conditions (EV)		x
Adopt overtaking position (EV)	x	
Assess gap ahead of LV (EV)		x
Driving free (OV)	x	
Keep in lane (LV)	x	x
Keep in lane (EV)	x	
Recheck road ahead (EV)		x
Abandon manoeuvre (EV)		x
Assess any new info for safety of manoeuvre again (EV)	x	x
Respond to EV's passing problems (LV)	x	
Adjust to adequate speed (LV)	x	
Respond to EV's passing problems (RV)	x	
Merge back into starting lane (EV)	x	x
Assess situation to enter safely (EV)	x	x
Continue observing road ahead (EV)		x
Driving free (LV)	x	
Keep in lane (OV)	x	
Respond to EV's passing problems (OV)	x	
Assess availability of safety margin in case of abort (EV)		x
Re-recheck road ahead (EV)		x
React to EV's overtaking (RV)	x	
Anticipate course of LV (EV)		x
Recognise that EV is experiencing problems passing (LV)	x	
Increase speed (EV)	x	
Assess overtaking opportunity again (EV)		x
Assess any new info for safety of manoeuvre (EV)	x	x
Watch for hazards located at roadside environment (EV)		x

References

1. Hughes, B.P.; Anund, A.; Falkmer, T. A comprehensive conceptual framework for road safety strategies. *Accid. Anal. Prev.* **2016**, *90*, 13–28. [CrossRef] [PubMed]
2. World Health Organization. *Global Status Report on Road Safety*; WHO Library Cataloguing-in-Publication Data; WHO: Geneva, Switzerland, 2020.

3. SAE On-Road Automated Vehicle Standards Committee. *Taxonomy and Definitions for Terms Related to on-Road Motor Vehicle Automated driving Systems*; J3016_201806; SAE International: Warrendale, PA, USA, 2018; pp. 1–16.
4. Hendricks, D.L.; Fell, J.C.; Freedman, M. *The Relative Frequency of Unsafe Driving Acts in Serious Injury Accidents*; Final report submitted to NHTSA under contract No. DOT NH 22 94 C 05020; Veridian engineering; Springer: Buffalo, NY, USA, 2001.
5. Otte, D.; Pund, B.; Jänsch, M. A new approach of accident causation analysis by seven steps ACASS. In Proceedings of the International Technical Conference on the Enhanced Safety of Vehicles, Stuttgart, Germany, 15–18 June 2009; National Highway Traffic Safety Administration: Washington, DC, USA, 2009; Volume 2009.
6. Singh, S. *Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey*; (No. DOT HS 812 115); NHTSA's National Center for Statistics and Analysis: Washington, DC, USA, 2015.
7. Dingus, T.A.; Guo, F.; Lee, S.; Antin, J.F.; Perez, M.; Buchanan-King, M.; Hankey, J. Driver crash risk factors and prevalence evaluation using naturalistic driving data. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 2636–2641. [CrossRef] [PubMed]
8. Woods, D.; Dekker, S. Anticipating the effects of technological change: A new era of dynamics for human factors. *Theor. Issues Ergon. Sci.* **2000**, *1*, 272–282. [CrossRef]
9. Drösler, J. Zur Methodik der Verkehrspsychologie. In *Psychologie des Straßenwesens*; Hoyos, C., Ed.; Huber: Bern, Switzerland; Stuttgart, Germany, 1965.
10. Rasmussen, J. Risk management in a dynamic society: A modelling problem. *Saf. Sci.* **1997**, *27*, 183–213. [CrossRef]
11. Awal Street Journal. Systems Thinking Speech by Dr. Russell Ackoff [Video]. Available online: https://www.youtube.com/watch?v=EbLh7rZ3rhU (accessed on 22 August 2021).
12. Grabbe, N.; Kellnberger, A.; Aydin, B.; Bengler, K. Safety of automated driving: The need for a systems approach and application of the functional resonance analysis method. *Saf. Sci.* **2020**, *126*, 104665. [CrossRef]
13. Bengler, K.; Winner, H.; Wachenfeld, W. No Human–No Cry? *Automatisierungstechnik* **2017**, *65*, 471–476. [CrossRef]
14. Wiener, E.L.; Curry, R.E. Flight-deck automation: Promises and problems. *Ergonomics* **1980**, *23*, 995–1011. [CrossRef]
15. Billings, C. *Aviation Automation*; Lawrence Erlbaum: Mahwah, NJ, USA, 1993.
16. Stanton, N.A.; Marsden, P. From fly-by-wire to drive-by-wire: Safety implications of automation in vehicles. *Saf. Sci.* **1996**, *24*, 35–49. [CrossRef]
17. Sarter, N.B.; Amalberti, R. (Eds.) *Cognitive Engineering in the Aviation Domain*; CRC Press: Boca Raton, FL, USA, 2000.
18. Noy, I.Y.; Shinar, D.; Horrey, W.J. Automated driving: Safety blind spots. *Saf. Sci.* **2018**, *102*, 68–78. [CrossRef]
19. Hollnagel, E. Understanding accidents-from root causes to performance variability. In Proceedings of the IEEE 7th conference on human factors and power plants, Scottsdale, AZ, USA, 19 September 2002; IEEE: New York, NY, USA, 2002; p. 1.
20. Qureshi, Z.H. A review of accident modelling approaches for complex socio-technical systems. In Proceedings of the 1757 twelfth Australian Workshop on Safety Critical Systems and Software and Safety-Related Programmable Systems, Adeleide, Australia, 30–31 August 2007; Australian Computer Society, Inc.: Darlinghurst, Australia, 2007; Volume 86, pp. 47–59.
21. Hollnagel, E. *Barriers and Accident Prevention Ashgate*; Routledge: Hampshire, UK, 2004.
22. Wienen, H.C.A.; Bukhsh, F.A.; Vriezekolk, E.; Wieringa, R.J. Accident Analysis Methods and Models—A Systematic Literature Review. 2017. Available online: https://ris.utwente.nl/ws/portalfiles/portal/13726744/Accident_Analysis_Methods_and_Models_a_Systematic_Literature_Review.pdf (accessed on 16 December 2021).
23. Salmon, P.M.; McClure, R.; Stanton, N.A. Road transport in drift? Applying contemporary systems thinking to road safety. *Saf. Sci.* **2012**, *50*, 1829–1838. [CrossRef]
24. Dekker, S.; Cilliers, P.; Hofmeyr, J.H. The complexity of failure: Implications of complexity theory for safety investigations. *Saf. Sci.* **2011**, *49*, 939–945. [CrossRef]
25. Perrow, C. *Normal Accidents: Living with High-Risk Technologies*; Basic Books: New York, NY, USA, 1984.
26. Larsson, P.; Dekker, S.W.; Tingvall, C. The need for a systems theory approach to road safety. *Saf. Sci.* **2010**, *48*, 1167–1174. [CrossRef]
27. Hughes, B.P.; Newstead, S.; Anund, A.; Shu, C.C.; Falkmer, T. A review of models relevant to road safety. *Accid. Anal. Prev.* **2015**, *74*, 250–270. [CrossRef] [PubMed]
28. Busch, C. *If You Can't Measure It—Maybe You Shouldn't: Reflections on Measuring Safety, Indicators, and Goals*; Mind The Risk: Wroclaw, Poland, 2019.
29. Hollnagel, E. *Safety-I and Safety-II: The Past and Future of Safety Management*; CRC Press: Boca Raton, FL, USA, 2014.
30. Dekker, S. *Drift into Failure: From Hunting Broken Components to Understanding Complex Systems*; CRC Press: Boca Raton, FL, USA, 2011.
31. Patriarca, R. Developing Risk and Safety Management Methods for Complex Sociotechnical Systems: From Newtonian Reasoning to Resilience Engineering. Ph.D. Thesis, Sapienza University of Rome, Rome, Italy, 2017. Available online: http://hdl.handle.net/11573/1194043 (accessed on 30 January 2021).
32. Hollnagel, E.; Woods, D.D.; Leveson, N. (Eds.) *Resilience Engineering: Concepts and Precepts*; Ashgate Publishing, Ltd.: Farnham, UK, 2006.
33. Grabbe, N.; Höcher, M.; Thanos, A.; Bengler, K. Safety Enhancement by Automated Driving: What are the Relevant Scenarios? *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2020**, *64*, 1686–1690. [CrossRef]
34. Hollnagel, E. Advancing resilient performance: From instrumental applications to second-order solutions. In Proceedings of the REA Symposium on Resilience Engineering Embracing Resilience, Toulouse, France, 21–24 June 2019.

35. Ferreira, P.N.; Cañas, J.J. Assessing operational impacts of automation using functional resonance analysis method. *Cogn. Technol. Work.* **2019**, *21*, 535–552. [CrossRef]
36. Nemeth, C. Erik Hollnagel: FRAM: The functional resonance analysis method, modeling complex socio-technical systems. *Cogn. Technol. Work.* **2013**, *1*, 117–118. [CrossRef]
37. Hollnagel, E. *FRAM: The Functional Resonance Analysis Method: Modelling Complex Socio-Technical Systems*; CRC Press: Boca Raton, FL, USA, 2012.
38. Woltjer, R.; Hollnagel, E. Functional modeling for risk assessment of automation in a changing air traffic management environment. In Proceedings of the 4th International Conference Working on Safety, Graz, Austria, 21–23 April 2008; Volume 30.
39. Patriarca, R.; Bergström, J. Modelling complexity in everyday operations: Functional resonance in maritime mooring at quay. *Cogn. Technol. Work.* **2017**, *19*, 711–729. [CrossRef]
40. Patriarca, R.; Di Gravio, G.; Woltjer, R.; Costantino, F.; Praetorius, G.; Ferreira, P.; Hollnagel, E. Framing the FRAM: A literature review on the functional resonance analysis method. *Saf. Sci.* **2020**, *129*, 104827. [CrossRef]
41. Hollnagel, E.; Pruchnicki, S.; Woltjer, R.; Etcher, S. Analysis of Comair flight 5191 with the functional resonance accident model. In Proceedings of the 8th International Symposium of the Australian Aviation Psychology Association, Sydney, Australia, 8–11 April 2008.
42. De Carvalho, P.V.R. The use of functional resonance analysis method (FRAM) in a mid-air collision to understand some characteristics of the air traffic management system resilience. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 1482–1498. [CrossRef]
43. Adriaensen, A.; Patriarca, R.; Smoker, A.; Bergström, J. A socio-technical analysis of functional properties in a joint cognitive system: A case study in an aircraft cockpit. *Ergonomics* **2019**, *62*, 1598–1616. [CrossRef]
44. Alm, H.; Woltjer, R. *Patient Safety Investigation through the Lens of FRAM. Human Factors: A System View of Human, Technology and Organization*; Shaker Publishing: Maastricht, The Netherlands, 2010; pp. 153–165.
45. Patriarca, R.; Falegnami, A.; Costantino, F.; Bilotta, F. Resilience engineering for socio-technical risk analysis: Application in neurosurgery. *Reliab. Eng. Syst. Saf.* **2018**, *180*, 321–335. [CrossRef]
46. Schutijser, B.C.F.M.; Jongerden, I.P.; Klopotowska, J.E.; Portegijs, S.; de Bruijne, M.C.; Wagner, C. Double checking injectable medication administration: Does the protocol fit clinical practice? *Saf. Sci.* **2019**, *118*, 853–860. [CrossRef]
47. Lundblad, K.; Speziali, J.; Woltjer, R.; Lundberg, J. FRAM as a risk assessment method for nuclear fuel transportation. In Proceedings of the 4th International Conference Working on Safety, Graz, Austria, 21–23 April 2008; Volume 1, pp. S223–S231.
48. Hollnagel, E.; Fujita, Y. The Fukushima disaster–systemic failures as the lack of resilience. *Nucl. Eng. Technol.* **2013**, *45*, 13–20. [CrossRef]
49. Macchi, L.; Oedewald, P.; Eitrheim, M.R.; Axelsson, C. Understanding maintenance activities in a macrocognitive work system. In Proceedings of the 30th European Conference on Cognitive Ergonomics, Edinburgh, UK, 28–31 August 2012; pp. 52–57.
50. Shirali, G.A.; Ebrahipour, V.; Mohammd Salahi, L. Proactive risk assessment to identify emergent risks using functional resonance analysis method (fram): A case study in an oil process unit. *Iran Occup. Health* **2013**, *10*, 33–46.
51. Franca, J.E.; Hollnagel, E.; dos Santos, I.J.L.; Haddad, A.N. Analysing human factors and non-technical skills in offshore drilling operations using FRAM (functional resonance analysis method). *Cogn. Technol. Work.* **2021**, *23*, 553–566. [CrossRef]
52. Praetorius, G.; Lundh, M.; Lützhöft, M. Learning from the past for proactivity: A re-analysis of the accident of the MV Herald of free enterprise. In Proceedings of the Fourth Resilience Engineering Symposium, Sophia-Antipolis, France, 8–10 June 2011; pp. 217–225.
53. Smith, D.; Veitch, B.; Khan, F.; Taylor, R. Using the FRAM to understand Arctic ship navigation: Assessing work processes during the Exxon Valdez grounding. *TransNav Int. J. Mar. Navig. Saf. Sea Transp.* **2018**, *12*, 447–457. [CrossRef]
54. Steen, R.; Aven, T. A risk perspective suitable for resilience engineering. *Saf. Sci.* **2011**, *49*, 292–297. [CrossRef]
55. Belmonte, F.; Schön, W.; Heurley, L.; Capel, R. Interdisciplinary safety analysis of complex socio-technological systems based on the functional resonance accident model: An application to railway traffic supervision. *Reliab. Eng. Syst. Saf.* **2011**, *96*, 237–249. [CrossRef]
56. Hlaing, K.P.; Aung, N.T.T.; Hlaing, S.Z.; Ochimizu, K. Functional resonance analysis method on road accidents in myanmar. In Proceedings of the 2nd International Conference on Advanced Information Technologies (ICAIT), Yangon, Myanmar, 1–2 November 2018; pp. 107–113.
57. Hirose, T.; Sawaragi, T.; Nomoto, H.; Michiura, Y. Functional safety analysis of SAE conditional driving automation in time-critical situations and proposals for its feasibility. *Cogn. Technol. Work.* **2021**, *23*, 639–657. [CrossRef]
58. Anfara, V.A.; Brown, K.M.; Mangione, T.L. Qualitative analysis on stage: Making the research process more public. *Educ. Res.* **2002**, *31*, 28–38. [CrossRef]
59. Creswell, J.W.; Miller, D.L. Determining validity in qualitative inquiry. *Theory Pract.* **2000**, *39*, 124–130. [CrossRef]
60. Creswell, J.W.; Poth, C.N. *Qualitative Inquiry and Research Design: Choosing among Five Approaches*; Sage Publications: Thousand Oaks, CA, USA, 2016.
61. Destatis. Verkehrsunfälle—Fachserie 8 Reihe 7—018. 2019. Available online: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Verkehrsunfaelle/Publikationen/Downloads-Verkehrsunfaelle/verkehrsunfaelle-jahr-2080700187004.html (accessed on 16 December 2021).
62. Richter, T.; Ruhl, S. *Untersuchung von Maßnahmen zur Prävention von Überholunfällen auf einbahnigen Landstraßen*; Unfallforschung der Versicherer, GDV: Berlin, Germany, 2014.

63. Netzer, M. Der Überholvorgang auf zweispurigen Straßen und seine Grundelemente unter besonderer Berücksichtigung der Verkehrssicherheit. 1966.
64. Walker, G.H.; Stanton, N.A.; Salmon, P.M. *Human Factors in Automotive Engineering and Technology*; Ashgate: Farnham, UK, 2015.
65. McKnight, A.J.; Adams, B.B. *Driver Education Task Analysis. Task Descriptions*; Human Resources Research Organization: Alexandria, VA, USA, 1970; Volume 1.
66. Hollnagel, E.; Hill, R. Instructions for use of the FRAM Model Visualiser (FMV). 2020. Available online: https://functionalresonance.com/onewebmedia/FMV_instructions_2.1.pdf (accessed on 1 March 2021).
67. Hollnagel, E. FRAM Model Interpreter. 2020. Available online: https://functionalresonance.com/onewebmedia/FMI%20basicPlus%20V3.pdf (accessed on 1 March 2021).
68. Risser, R.; Brandstätter, C. Die Wiener Fahrprobe. Freie Beobachtung. 1985, Volume 21. Available online: https://trid.trb.org/view/1034307 (accessed on 16 December 2021).
69. Patriarca, R.; Di Gravio, G.; Costantino, F. A Monte Carlo evolution of the functional resonance analysis method (FRAM) to assess performance variability in complex systems. *Saf. Sci.* **2017**, *91*, 49–60. [CrossRef]
70. Patriarca, R.; Di Gravio, G.; Costantino, F. myFRAM: An open tool support for the functional resonance analysis method. In Proceedings of the 2017 2nd International Conference on System Reliability and Safety (ICSRS), Milan, Italy, 20–22 December 2017; IEEE: New York, NY, USA, 2017; pp. 439–443.
71. Falegnami, A.; Costantino, F.; Di Gravio, G.; Patriarca, R. Unveil key functions in socio-technical systems: Mapping FRAM into a multilayer network. *Cogn. Technol. Work.* **2019**, *22*, 877–899. [CrossRef]
72. Bellini, E.; Ceravolo, P.; Nesi, P. Quantify resilience enhancement of UTS through exploiting connected community and internet of everything emerging technologies. *ACM Trans. Internet Technol. (TOIT)* **2017**, *18*, 1–34. [CrossRef]
73. Borgatti, S.P. Centrality and network flow. *Soc. Netw.* **2005**, *27*, 55–71. [CrossRef]
74. Patriarca, R.; Bergström, J.; Di Gravio, G. Defining the functional resonance analysis space: Combining Abstraction Hierarchy and FRAM. *Reliab. Eng. Syst. Saf.* **2017**, *165*, 34–46. [CrossRef]
75. Rasmussen, J.; Lind, M. *Coping with Complexity*; Risø National Laboratory: Roskilde, Denmark, 1981.
76. Parasuraman, R.; Sheridan, T.B.; Wickens, C.D. A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2000**, *30*, 286–297. [CrossRef] [PubMed]
77. Fitts, P.M. *Human Engineering for an Effective Air-Navigation and Traffic-Control System*; National Research Council: Washington, DC, USA, 1951.
78. Sheridan, T.B. *Telerobotics, Automation, and Human Supervisory Control*; MIT Press: Cambridge, MA, USA, 1992.
79. Inagaki, T. Adaptive automation: Sharing and trading of control. In *Handbook of Cognitive Task Design*; CRC Press: Boca Raton, FL, USA, 2003; Chapter 8; pp. 147–169.
80. Hollnagel, E.; Woods, D.D. *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*; CRC Press: Boca Raton, FL, USA, 2005.
81. Hollnagel, E. A function-centred approach to joint driver-vehicle system design. *Cogn. Technol. Work.* **2006**, *8*, 169–173. [CrossRef]
82. Zhang, Y.; Angell, L.; Bao, S. A fallback mechanism or a commander? A discussion about the role and skill needs of future drivers within partially automated vehicles. *Transp. Res. Interdiscip. Perspect.* **2021**, *9*, 100337. [CrossRef]
83. Michon, J.A. A critical view of driver behavior models: What do we know, what should we do? In *Human Behavior and Traffic Safety*; Springer: Boston, MA, USA, 1985; pp. 485–524.
84. Kauer, M.; Schreiber, M.; Bruder, R. How to conduct a car? A design example for maneuver based driver-vehicle interaction. In Proceedings of the 2010 IEEE Intelligent Vehicles Symposium, La Jolla, CA, USA, 21–24 June 2010; IEEE: New York, NY, USA, 2010; pp. 1214–1221.
85. Franz, B.; Kauer, M.; Bruder, R.; Geyer, S. pieDrive—A new driver-vehicle interaction concept for maneuver-based driving. In Proceedings of the 2012 International IEEE Intelligent Vehicles Symposium Workshops (W2: Workshop on Human Factors in Intelligent Vehicles), Alcala de Henares, Spain, 3–7 June 2012; Toledo-Moreo, R., Bergasa, L.M., Sotelo, M.A., Eds.; IEEE: New York, NY, USA, 2012.
86. Walch, M.; Sieber, T.; Hock, P.; Baumann, M.; Weber, M. Towards cooperative driving: Involving the driver in an autonomous vehicle's decision making. In Proceedings of the 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Ann Arbor, MI, USA, 24–26 October 2016; pp. 261–268.
87. Walch, M.; Woide, M.; Mühl, K.; Baumann, M.; Weber, M. Cooperative overtaking: Overcoming automated vehicles' obstructed sensor range via driver help. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Utrecht, The Netherlands, 21–25 September 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 144–155.
88. Flemisch, F.; Abbink, D.A.; Itoh, M.; Pacaux-Lemoine, M.P.; Weßel, G. Joining the blunt and the pointy end of the spear: Towards a common framework of joint action, human–machine cooperation, cooperative guidance and control, shared, traded and supervisory control. *Cogn. Technol. Work.* **2019**, *21*, 555–568. [CrossRef]
89. Pacaux-Lemoine, M.P.; Flemisch, F. Layers of shared and cooperative control, assistance, and automation. *Cogn. Technol. Work.* **2019**, *21*, 579–591. [CrossRef]
90. Inagaki, T. Traffic systems as joint cognitive systems: Issues to be solved for realizing human-technology coagency. *Cogn. Technol. Work.* **2010**, *12*, 153–162. [CrossRef]

91. Inagaki, T.; Sheridan, T.B. A critique of the SAE conditional driving automation definition, and analyses of options for improvement. *Cogn. Technol. Work.* **2019**, *21*, 569–578. [CrossRef]
92. Wickens, C.D. Designing for situation awareness and trust in automation. *IFAC Proc. Vol.* **1995**, *28*, 365–370. [CrossRef]
93. Endsley, M.R.; Kiris, E.O. The out-of-the-loop performance problem and level of control in automation. *Hum. Factors* **1995**, *37*, 381–394. [CrossRef]
94. Parasuraman, R.; Riley, V. Humans and automation: Use, misuse, disuse, abuse. *Hum. Factors* **1997**, *39*, 230–253. [CrossRef]
95. Sarter, N.B.; Woods, D.D.; Billings, C.E. Automation surprises. In *Handbook of Human Factors and Ergonomics*; Wiley: New York, NY, USA, 1997; pp. 1926–1943.
96. Inagaki, T.; Stahre, J. Human supervision and control in engineering and music: Similarities, dissimilarities, and their implications. *Proc. IEEE* **2004**, *92*, 589–600. [CrossRef]
97. Wilde, G.J. The theory of risk homeostasis: Implications for safety and health. *Risk Anal.* **1982**, *2*, 209–225. [CrossRef]
98. Hollnagel, E. *From Function Allocation to Function Congruence. Coping with Computers in the Cockpit (A 00-40958 11-54)*; Ashgate Publishing: Aldershot, UK; Brookfield, VT, USA, 1999; pp. 29–53.
99. MacKinnon, R.J.; Pukk-Härenstam, K.; Kennedy, C.; Hollnagel, E.; Slater, D. A novel approach to explore Safety-I and Safety-II perspectives in in situ simulations—The structured what if functional resonance analysis methodology. *Adv. Simul.* **2021**, *6*, 21. [CrossRef]
100. Hill, R.; Boult, M.; Sujan, M.; Hollnagel, E.; Slater, D. Predictive Analysis of Complex Systems' Behaviour. 2020. Available online: https://www.researchgate.net/profile/David-Slater/publication/343944100_PREDICTIVE_ANALYSIS_OF_COMPLEX_SYSTEMS\T1\textquoteright_BEHAVIOUR_SWIFTFRAM/links/5f4907e0299bf13c5047f8d3/PREDICTIVE-ANALYSIS-OF-COMPLEX-SYSTEMS-BEHAVIOUR-SWIFTFRAM.pdf (accessed on 16 December 2021).
101. Boggs, A.M.; Arvin, R.; Khattak, A.J. Exploring the who, what, when, where, and why of automated vehicle disengagements. *Accid. Anal. Prev.* **2020**, *136*, 105406. [CrossRef]
102. Dixit, V.V.; Chand, S.; Nair, D.J. Autonomous vehicles: Disengagements, accidents and reaction times. *PLoS ONE* **2016**, *11*, e0168054. [CrossRef]
103. Favarò, F.M.; Eurich, S.O.; Nader, N. Analysis of disengagements in autonomous vehicle technology. In Proceedings of the 2018 Annual Reliability and Maintainability Symposium (RAMS), Reno, NV, USA, 22–25 January 2018; IEEE: New York, NY, USA, 2018; pp. 1–7.
104. Durth, W.; Habermehl, K. *Überholvorgänge auf einbahnigen Straßen. Forschung Straßenbau und Straßenverkehrstechnik. Bundesminister für Verkehr*; Abt. Strassenbau: Bonn, Germany, 1986.
105. Gründl, M. Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial und die Gestaltung von Fahrerassistenzsystemen. Ph.D. Thesis, University of Regensburg, Regensburg, Germany, July 2005.
106. Winner, H. Quo vadis, FAS? In *Handbuch Fahrerassistenzsysteme*; Springer Vieweg: Wiesbaden, Germany, 2015; pp. 1167–1186.
107. Wachenfeld, W.; Winner, H. The release of autonomous vehicles. In *Autonomous Driving*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 425–449.
108. Riedmaier, S.; Ponn, T.; Ludwig, D.; Schick, B.; Diermeyer, F. Survey on scenario-based safety assessment of automated vehicles. *IEEE Access* **2020**, *8*, 87456–87477. [CrossRef]
109. Junietz, P.; Wachenfeld, W.; Klonecki, K.; Winner, H. Evaluation of different approaches to address safety validation of automated driving. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; IEEE: New York, NY, USA, 2018; pp. 491–496.
110. Bridges, K.E.; Corballis, P.M.; Hollnagel, E. "Failure-to-identify" hunting incidents: A resilience engineering approach. *Hum. Factors* **2018**, *60*, 141–159. [CrossRef]
111. *Road Vehicles—Functional Safety—Part 3: Concept Phase*; ISO Standard 26262. 2018. Available online: https://www.iso.org/standard/68385.html (accessed on 16 December 2021).
112. Underwood, P.; Waterson, P. A critical review of the STAMP, FRAM and Accimap systemic accident analysis models. In *Advances in Human Aspects of Road and Rail Transportation*; CRC Press: Boca Raton, FL, USA, 2012; pp. 385–394. Available online: https://www.researchgate.net/profile/Patrick-Waterson/publication/236023374_A_critical_review_of_the_STAMP_FRAM_and_Accimap_systemic_accident_analysis_models/links/5696572608ae1c42790399a1/A-critical-review-of-the-STAMP-FRAM-and-Accimap-systemic-accident-analysis-models.pdf (accessed on 16 December 2021).
113. Salmon, P.M.; Read, G.J.; Walker, G.H.; Stevens, N.J.; Hulme, A.; McLean, S.; Stanton, N.A. Methodological issues in systems Human Factors and Ergonomics: Perspectives on the research–practice gap, reliability and validity, and prediction. *Hum. Factors Ergon. Manuf. Serv. Ind.* **2020**, *32*, 6–19. [CrossRef]
114. Constantino, F.; Di Gravio, G.; Falegnami, A.; Patriarca, R.; Tronci, M.; De Nicola, A.; Vicoli, G.; Villani, M.L. Crowd sensitive indicators for proactive safety management: A theoretical framework. In Proceedings of the 30th European Safety and Reliability Conference ESREL and 15th Probabilistic Safety Assessment and Management Conference; PSAM 15 2020. Research Publishing Services: Singapore, 2020; pp. 1453–1458. [CrossRef]
115. Farooqi, A.; Ryan, B.; Cobb, S. Using expert perspectives to explore factors affecting choice of methods in safety analysis. *Saf. Sci.* **2022**, *146*, 105571. [CrossRef]
116. Maurer, M.; Gerdes, J.C.; Lenz, B.; Winner, H. *Autonomous Driving*; Springer: Berlin/Heidelberg, Germany, 2016; Volume 10, pp. 978–983.

117. Hollnagel, E. *Synesis: The Unification of Productivity, Quality, Safety and Reliability*; Routledge: Milton Park, UK, 2020.
118. Patriarca, R.; Del Pinto, G.; Di Gravio, G.; Constantino, F. FRAM for systemic accident analysis: A matrix representation of functional resonance. *Int. J. Reliab. Qual. Saf. Eng.* **2018**, *25*, 1850001.

E Article 4: “Assessing the reliability and validity of an FRAM model: the case of driving in an overtaking scenario”

Grabbe, N., Arifagic, A., & Bengler, K. (2022). Assessing the reliability and validity of an FRAM model: the case of driving in an overtaking scenario. *Cognition, Technology & Work, 24(3),* 483-508. https://doi.org/10.1007/s10111-022-00701-7

Cognition, Technology & Work (2022) 24:483–508
https://doi.org/10.1007/s10111-022-00701-7

ORIGINAL ARTICLE

Assessing the reliability and validity of an FRAM model: the case of driving in an overtaking scenario

Niklas Grabbe[1] · Almin Arifagic[1] · Klaus Bengler[1]

Received: 21 February 2022 / Accepted: 2 May 2022 / Published online: 26 May 2022

Abstract

Over the past two decades, systemic-based risk assessment methods have garnered more attention, and their use and popularity are growing. In particular, the functional resonance analysis method (FRAM) is one of the most widely used systemic methods for risk assessment and accident analysis. FRAM has been progressively evolved since its starting point and is considered to be the most recent and promising step in understanding socio-technical systems. However, there is currently a lack of any formal testing of the reliability and validity of FRAM, something which applies to Human Factors and Ergonomics research as a whole, where validation is both a particularly challenging issue and an ongoing concern. Therefore, this paper aims to define a more formal approach to achieving and demonstrating the reliability and validity of an FRAM model, as well as to apply this formal approach partly to an existing FRAM model so as to prove its validity. At the same time, it hopes to evaluate the general applicability of this approach to potentially improve the performance and value of the FRAM method. Thus, a formal approach was derived by transferring both the general understanding and definitions of reliability and validity as well as concrete methods and techniques to the concept of FRAM. Consequently, predictive validity, which is the highest maxim of validation, was assessed for a specific FRAM model in a driving simulator study using the signal detection theory. The results showed that the predictive validity of the FRAM model is limited and a generalisation with changing system conditions is impossible without some adaptations of the model. The applicability of the approach is diminished because of several methodological limitations. Therefore, the reliability and validity framework can be utilised to calibrate rather than validate an FRAM model.

Keywords FRAM · Validation · Driving · Overtaking manoeuvre

1 Introduction

Risk assessment is a crucial aspect of Human Factors and Ergonomics (HFE) research. Instead of the reactive approach taken in accident analyses, which looks at a particular erroneous scenario, risk assessment adopts a proactive approach, trying to identify hazards or looking for what could happen in the future to prevent or mitigate adverse events or to facilitate desirable outcomes. Over the past 20 years, systemic based risk assessment methods have garnered more attention and their use and popularity are growing (e.g., Dallat et al. 2017; Hollnagel 2012; Hughes et al. 2015; Hulme et al. 2019; Larsson et al. 2010; Leveson 2004; Salmon et al. 2012). These methods try to describe performance at the level of the overall system and see the accident process as a complex and interwoven event that cannot be broken down into its individual parts. Emerging events caused by complex and non-linear interactions between the various system parts can affect the performance of the system and cause an accident (Laaraj and Jawab 2018; Qureshi 2007; Wienen et al. 2017). In general, systemic models acknowledge the complexity and socio-technical nature of systems, and further emphasise the need for an understanding of the functional abstraction of the system, rather than structural decomposition (Rasmussen 1997).

In particular, the functional resonance analysis method (FRAM) (Hollnagel 2012) is one of the most widely used systemic methods for risk assessment and accident analysis. It allows the modelling of mechanisms within complex socio-technical systems (STS), including their interfaces between humans and technology, coupling and dependency

✉ Niklas Grabbe
n.grabbe@tum.de

[1] Chair of Ergonomics, Technical University of Munich, Boltzmannstr. 15, 85748 Garching, Germany

effects, nonlinear interactions between elements, and functional variability (Woltjer and Hollnagel 2008). In general, the results of an FRAM analysis contribute to an understanding of real work and reveal unsafe functional interactions within one agent and between different agents; these are needed to assist risk management as regards the proactive assessment of technological changes and their impacts (Ferreira and Cañas 2019; Patriarca and Bergström 2017). In addition, FRAM should form the basis for systemic risk assessments in complex STS, for example for contemporary applications, such as automated driving in road traffic (Grabbe et al. 2020, 2022). These authors do so by providing a useful understanding of the actual system mechanisms and interactions that are needed to assist the system design, enhanced by considering non-linear, complex, and emergent system behaviour (Grabbe et al. 2022). In the past, FRAM has been widely used and enhanced methodologically in a variety of domains for retrospective as well as prospective analyses, as detailed in a comprehensive review by Patriarca et al. (2020). Hence, FRAM has been progressively evolved since its starting point in 2004 (Hollnagel 2004) and is considered to be the most recent and promising step in understanding STS (Nemeth 2013).

However, there is currently a lack of any formal testing of the reliability and validity of FRAM. This applies to the HFE research as a whole, where validation is both a particularly challenging issue and an ongoing concern (Stanton and Young 1999a, 2003; Stanton 2016). In fact, Stanton and Young (1999a) stated that practitioners often assume validity, but seldom test and prove it empirically. Furthermore, methods are often chosen by practitioners that are based on familiarity and ease of use rather than on reliability and validity evidence (Stanton et al. 2013). Thus, findings from the application of HFE methods suffer from an objective evaluation, making the research findings questionable. However, HFE methods must prove that these methods can intentionally work in their applied domains (Stanton 2014) and to promote the credibility of HFE methods and their whole community (Stanton 2016). In this context, and since FRAM should form the basis for systemic risk assessments in complex STS (Grabbe et al. 2020, 2022), validation is an absolute priority and a compulsory aspect in engineering disciplines (where HFE is part of it), especially in the aforementioned field of automated driving, due to the enormous societal impact which benefits FRAM by providing a clear evaluation of its performance and value.

Thus, this paper aims to first define a more formal approach to achieving and demonstrating the reliability and validity of an FRAM model that forms the basis for risk identification and design recommendations within the FRAM method, and second, to apply this formal approach partly to an existing FRAM model so as to prove its validity, and to evaluate the general applicability of this approach.

The remainder of this paper is structured as follows. Section 1.1 summarises the theoretical foundations and individual analytical steps of FRAM, as well as previous validation approaches. Following on from this, Sect. 1.2 summarises approaches for testing the reliability and validity of HFE methods. Section 2 outlines the understanding and definitions of reliability and validity in literature and transfers these to the context of FRAM to define a framework that addresses the reliability and validity of FRAM models. In Sect. 3, we describe the methodology for the evaluation of predictive validity in a driving simulator experiment. Section 4 presents the results, including the evaluation of the predictive validity of the analysed FRAM model according to the three different research questions of the study. Section 5 then discusses the results with respect to the research goals of this paper and also outlines methodological limitations. Finally, a brief conclusion and outlook for future research are provided in Sect. 6.

1.1 Basics of FRAM and previous validation approaches

The purpose of the model produced by the FRAM method is to describe and understand what is happening in an STS in terms of functions rather than components. An FRAM model focuses on adjustments to everyday performance, which usually contribute to things going right. In rare cases, these performance adjustments aggregate in unexpected ways, leading to functional resonance, with accidents being the most extreme result.

FRAM relies on four principles (the equivalence of success and failures, approximate adjustments, emergence, and functional resonance), and follows four steps (modelling the system by identifying its functions, identifying the function's performance variability, aggregating the variability, and managing the variability), as detailed in Hollnagel (2012). The steps are briefly described in the following. In the first step, the essential functions of a system are identified to build a model. Basically, each function is characterised by six aspects (i.e., input, output, precondition, resource, control, and time), which couple each function with several other functions representing a specific instantiation of the model that traditionally is represented graphically by hexagons. Furthermore, the functions can be divided into two classes: foreground and background functions. Foreground functions are the core of the analysis and may vary significantly during an instantiation of the model. In contrast, background functions are stable and represent common conditions as a system boundary that are used by foreground functions. The second step is to specify the performance variability of each function that can be characterised in its simple form using two phenotypes, namely, timing and precision. Here, the function's output in terms of timing can occur too early, on time, too late, or not at all, whereas for precision, the output can be precise, acceptable, or imprecise (Hollnagel 2012). In the third step, the variability is aggregated to understand how the variability can propagate through the system and where functional resonance emerges leading to adverse outcomes. This is done by defining upstream–downstream couplings, where variability can be caused through couplings of upstream functions, when the output used as input or resource, for example, is variable and thus affects the variability of downstream functions. The fourth and final step consists of the monitoring and management of the previously identified performance variability to ensure the safety and performance of the system.

In the past, some attempts were made to formally verify an FRAM model. The first attempt at formal verification was the FRAM model-based safety assessment that used model checking and theorem proving to verify the FRAM model so as to determine whether pre-set safety requirements can be observed (Yang and Tian 2015). The same authors enhanced this approach using the Simple Promela Interpreter (SPIN) tool and applied it to develop an air traffic management system. The analysis demonstrated that FRAM can benefit from a formal verification with the aid of model checking through more rigorous computation that improves its efficiency and accuracy (Yang et al. 2017). In addition, the software tool FRAM Model Interpreter (FMI) (Hollnagel 2020) has recently become available, which is a stepwise automatic interpretation of the syntactical and logical correctness of an FRAM model to formally check and adjust its consistency and completeness. With regard to validation, subjective evaluation through interviews with experts, workshops, and discussions was mainly used to improve the face validity of developed FRAM models, as pointed out by Bridges et al. (2018), Kaya et al. (2019), and Ross et al. (2018). The reason may be associated with an experts' deep knowledge of the work system and daily operations, which can help to enrich developed FRAM models and to provide more reliable models (Salehi et al. 2021). However, a more formal approach for validation is still lacking.

1.2 Previous approaches to testing the reliability and validity of HFE methods

On the whole, studies are rarely conducted that report the reliability or validity of HFE methods. However, some examples can be found and are summarised in the following. The reliability of ergonomics methods is often assessed using a test–retest paradigm (Baysari et al. 2011). Examples of the measures used here include percentage agreement (Baber and Stanton 1996; Baysari et al. 2011; O'Connor 2008), Pearson’s correlation (Harris et al. 2005; Stanton and Young 2003), the index of concordance (e.g., Olsen and Shorrock 2010), and Cohen’s kappa (e.g., Makeham et al. 2008).

Studies assessing the validity of ergonomics methods can also be found in literature (Baber and Stanton 1996; Stanton et al. 2009; Stanton and Young 2003). Many of these have focussed on human reliability and error prediction methods in general (Baysari et al. 2011; Kirwan et al. 1997; Stanton and Young 2003) or more specifically on the systematic human error reduction and prediction approach (SHERPA) (Stanton and Stevenage 1998) and task analysis for error identification (TAFEI) (Stanton and Baber 2005). In these studies, the validity of methods was assessed by comparing a method’s results (e.g., errors predicted) against actual observations (e.g., errors observed). More recently, system analysis methods, such as the cognitive work analysis (Cornelissen et al. 2014), a factor classification scheme for Rasmussen's Accimap (Goode et al. 2017), the networked hazard analysis and risk management system (Net-HARMS) (Hulme et al. 2021a), and the operator event sequence diagrams (Stanton et al. 2021a, b, c, d) have also been empirically validated. Furthermore, there has been a thorough comparison of intra-rater reliability and criterion-referenced concurrent validity between three systems-based risk assessment approaches: the systems-theoretic process analysis (STPA) method, the event analysis of systemic teamwork broken links (EAST-BL) method, and the Net-HARMS method (Hulme et al. 2021b; see also Hulme et al. 2021c). In general, quantitative methods to compare expert results versus novice results (or predicted versus actual outcomes) are often based on the use of signal detection theory (SDT) to calculate the sensitivity of the method under analysis (Baber and Stanton 1994; Stanton et al. 2009; Stanton and Young 2003). The SDT and its metrics are commonly used to assess the reliability and validity of ergonomics methods, such as human error prediction (Stanton et al. 2009). This was pioneered in particular by Stanton and Young (1999a, b) as a means of establishing empirical validity of methods.

A comparison of the reliability and validity of a range of HFE methods has been undertaken by Stanton and Young (1999a, b, 2003), which showed that the methods vary quite considerably in their performance. This demonstrates the urgent need for more reliability and validation studies of other HFE methods, and in particular FRAM. Moreover, FRAM follows a safety-II perspective (Hollnagel 2014) for which validity is seldomly addressed instead of safety-I (Hollnagel 2014) based methods as, e.g., human error analysis methods as mentioned previously.

2 Proposed reliability and validity framework

2.1 Understanding and definitions of reliability and validity

According to Stanton and Young (1999a), reliability and validity are interrelated, where a method can only be valid if it is reliable but may be reliable and not valid. Thus, these two criteria have to be evaluated mutually.

Reliability is a measure of the stability of the method over time and across analysts, ideally demonstrating that the application of an ergonomics method will result in the same results if it is used by different people (inter-rater) or at different points in time by the same people (intra-rater) (Stanton et al. 2016). This is often assessed using a test–retest paradigm between experts and novices, including measures, such as percentage agreement and Cohen's Kappa (e.g., Baysari et al. 2011; Hulme et al. 2021a; Makeham et al. 2008).

When considering validity, we have to distinguish between the following two main terms: verification and validation. According to Balci (1998), verification determines whether the formal implementation of a model is correct, which deals with building the model correctly. On the other hand, validation determines whether a model can be substituted for the real system for the intended purposes and objectives in the applied domain, which deals with building the right model. Overall, a model must be useful with regard to its objective, which means providing a reasonably accurate answer to the question to be answered (Liebl 2018, p. 203). Consequently, the concept of validity has to be guided by this requirement and should not be regarded as absolute (Schrank and Holt 1967). This has various implications for the nature of validation (Liebl 2018, pp. 203–205):

- *model-individual,* meaning that it is impossible to postulate a standardised validation procedure due to various forms and applications of models. Rather, the required validity criteria and their weighting change depending on the problem (Banks et al. 1987).
- *gradual,* showing how good or bad a model is in fulfilling its purpose and describing the validation process as a trade-off between additional costs/effort and the added information value of increased validity (Van Horn 1971).
- *result of a negotiation process,* according to which the validity of a model largely equates to the question of credibility and acceptance. Within this process, it is negotiated when the model is considered sufficiently valid and which validity criteria and methods should be applied (cf. Sargent 1984).
- *continuous and iterative,* meaning that validation takes place during the entire development process and "confidence is built into the model as the study proceeds" (Bulgren 1982, p. 126) rather than depicting a separate section at the end as an end state.

Furthermore, different categories of validation can be found in literature. For instance, Liebl (2018) distinguishes between outcome-based, function-based, and theory-based validation. Outcome-based validation aims to compare results, checking the extent to which the model produces results that match those of the real system. Function-based validation comes into play when the real system is not fully observable, so one has to validate exclusively on the model itself. Here, the reaction mode of the model is checked for plausibility, hence validity ultimately presents itself as a failed falsification of the model (Hanssmann 2018, p. 93). Theory-based validation compares the model results with theoretically expected results, which usually come from analytical models or literature.

As for HFE methods, Stanton and Young (1999b) proposed four types of validity for ergonomics methods: construct, content, concurrent and predictive. Construct validity concerns the underlying theoretical basis of a method. Content validity relates to the credibility that a method can achieve with its users, which can also be referred to as face validity. Finally, concurrent and predictive validity address the extent to which an analysed performance is representative of the performance that might have been analysed, where concurrent validity describes the current performance sampled, and predictive validity (i.e., criterion-referenced empirical validity) concerns the performance in the future. Furthermore, HFE methods should possess a certain level of concurrent or predictive validity suitable for their application (Stanton 2016). However, it is debatable as to whether all ergonomics methods have to fulfil all four types of validation, as shown by a distinction between analytic and evaluative methods, assuming that construct and content validity might be sufficient for analytic methods, whereas predictive validity might be required for evaluative methods (Annett 2002).

Finally, various concrete techniques can be used to test the aforementioned validation and verification types. Balci (1998, p. 355) presented an overview of more than 75 techniques, placing them into four categories: informal, static, dynamic, and formal. The use of mathematical and logic formalism by the techniques increases from informal to formal. Informal techniques are the most commonly used and rely heavily on subjectivity. Examples include audits, face validation, turing tests, and walkthroughs. Static techniques assess the model's accuracy based on the characteristics of the static model design, including, for example, control analysis, semantic and syntax analysis as well as structural analysis. Dynamic models, on the other hand, evaluate the model based on its execution behaviour, including, among others, predictive validation, sensitivity analysis, and statistical techniques. Last but not least, the formal techniques are quite objective and are based on a mathematical proof of correctness, for instance, induction and logical deduction.

2.2 Transfer and applicability to FRAM

As we have seen before, validity is not an absolute concept, but rather a relative one. Thus, there is no standard approach to validity. Instead, an approach to prove validity and reliability has to be developed for each method itself according to the features and context of the application. Therefore, the aforementioned knowledge will be transferred to the concept of FRAM to define one potential approach to demonstrate reliability and validity for an FRAM model in the following (see Fig. 1). It should be pointed out that we have to distinguish between an FRAM model and a particular instantiation of the model when trying to define a validation approach for the FRAM method. According to Hollnagel (2012), the functions are potentially coupled in an FRAM model, meaning that there is no predetermined a priori order or fixed sequence of the functions, whereby the functions actually become coupled in an instantiation for a specific set of conditions, resulting in temporal and causal relations. Against this background, validation is only possible for a particular instantiation of an FRAM model, but not for an FRAM model in general. For the sake of simplicity, we use the term "FRAM model" as meaning an "instantiation of an FRAM model" in this paper.

Basically, FRAM is a qualitative modelling method that offers great flexibility in terms of how it is applied and used, since it is a method-sine-model which means that FRAM is used as a method to produce a model and not vice versa (Hollnagel 2012, pp. 127–133). In addition, an experienced team of experts is required to analyse and model the system (Accou and Reniers 2019; Jensen and Aven 2018; Pereira 2013), where the quality of the output in FRAM directly depends on the team of experts and the information they provide as input for the functions and their variability (Salehi et al. 2021). Although some practical guidance material exists in Hollnagel et al. (2014), there is no explicit standard for determining how much information should be included in the analytical process to define the objective, scope, and granularity of the model, as highlighted by Anvarifar et al. (2017), Grabbe et al. (2020), Li et al. (2019), and Patriarca et al. (2017). Due to these low limitations or regulations regarding modelling, as well as the strong dependency between model outputs and the competence of the modeller team, an FRAM model is ultimately subject to a very strong subjective component. This means that when applied to the same work context and using the method traditionally, an FRAM model and its risk derivation are unlikely to be congruent between different users and even with the same user on a different occasion. For this reason, the classic test–retest

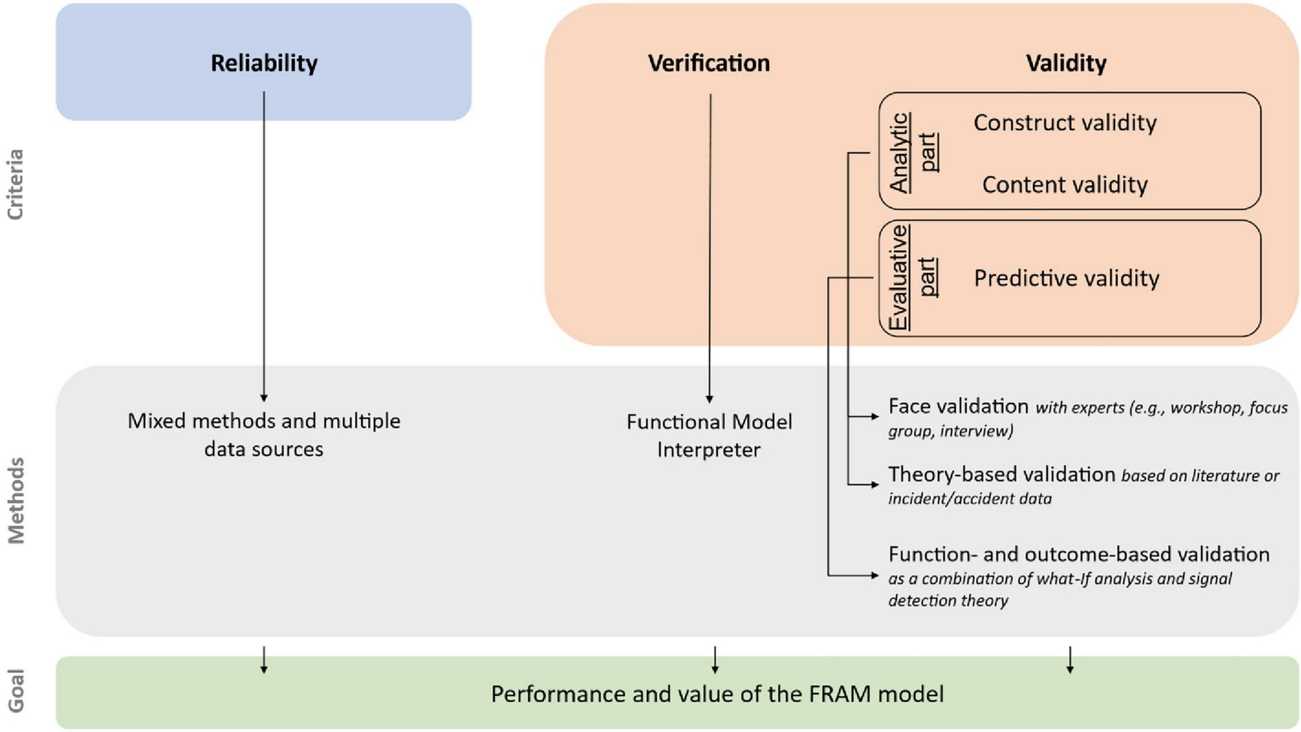

Criteria
Reliability
Verification
Validity
Analytic part
Construct validity
Content validity
Evaluative part
Predictive validity
Methods
Mixed methods and multiple data sources
Functional Model Interpreter
Face validation with experts (e.g., workshop, focus group, interview)
Theory-based validation based on literature or incident/accident data
Function- and outcome-based validation as a combination of what-If analysis and signal detection theory
Goal
Performance and value of the FRAM model

Fig. 1 Validation approach for an FRAM model

paradigm, which is often used to assess the reliability of HFE methods as mentioned in Sect. 2.1, seems inappropriate in the context of FRAM, particularly for largely complex FRAM models. Therefore, the reliability for an FRAM model cannot be proven but it can be achieved and increased using mixed methods and multiple data sources, such as document reviews, interviews, observations and simulations, as well as workshops and focus groups (see Fig. 1). These help to integrate multiple limited perspectives and dimensions that adhere to the verification strategies of Creswell and Miller (2000), complying with the four qualitative terms of credibility, transferability, dependability, and confirmability (Anfara et al. 2002) to improve the quality, scientific rigour, and trustworthiness of the model. One application of this can be found in Adriaensen et al. (2019) and Grabbe et al. (2022) in the context of aircraft cockpits and automated driving, respectively.

In addition, an FRAM model can be simply verified through the already established FMI (Hollnagel 2020) software that automatically interprets and parses the syntactical and logical correctness of an FRAM model step-by-step to formally check and adjust its structure with regard to consistency and completeness while obeying the FRAM “rules” (see Fig. 1). An important part of this is the identification of orphans or potential auto-loops as well as the question of whether relations between functions are mutually consistent, thus allowing an event to develop as intended. The use of FMI can be enhanced by other tools, such as model checking and theorem proving, as described in Sect. 1.1.

Validity should be divided into construct, content, and predictive validity according to Stanton and Young (1999b), since FRAM is an ergonomics method. In this case, concurrent validity is omitted, because an FRAM model does not generate absolute outputs; the outputs can only be evaluated relatively if something is changed in the model. This means that only future performance and not current performance can be validated. However, this is not a problem, since predictive validity is the higher maxim of the two anyway. Furthermore, FRAM is both an analytic and evaluative method. The analytic part is used through the qualitative and traditional application to gain an understanding of the mechanisms that underlie the functional interactions between system elements by modelling to comprehend what is happening, for example, to facilitate design decisions or to identify sources of performance failures and successes. In contrast to this, the evaluative part is used more in a semi-quantitative approach to measure and predict a certain parameter, such as performance variability, which is the fundamental factor explaining system behaviour in the FRAM method with its core principles of performance adjustments, emergence, and functional resonance. The analytic and evaluative parts are covered by construct and content validity, and predictive validity, respectively (cf. Annett 2002) (see Fig. 1). Construct validity should be ensured through the strong and sound systems theory basis of FRAM, as well as the tremendous credibility that the method gained amongst users over the last decade (cf. Patriarca et al 2020), which is also an argument for the content validity. Thus, construct validity can be generally assumed for an FRAM model as long as the method and its principles were correctly and comprehensively used, once again emphasising the strong dependency between an FRAM model’s output quality and the experience and training of the user and modeller as mentioned above. Content validity can mainly be proved by face validity using subjective evaluation through interviews, workshops, and discussions with experts who have a deep knowledge of normal work systems and daily operations, as already applied by Bridges et al. (2018), Kaya et al. (2019), and Ross et al. (2018). In addition, a theory-based validation could be used to further increase the content validity by comparing the FRAM model’s outputs with both other models or indicators in literature or incident and accident reports (including contributory factors and reasons) regarding the same application context. For instance, Bridges et al. (2018) modelled real accidents as "Mini FRAMs" based on accident reports that served as a comparison for the logic of the overall FRAM model.

Finally, predictive validity could be demonstrated by a mixture of function- and outcome-based validation. The reason for the combination is that an outcome-based validation alone is not possible, because an FRAM model does not generate absolute, observable outputs as a final product of the entire model. Instead, it must be linked to a function-based validation to produce relative, observable outputs through controlled variations in the model. The function-based validation can be realised by a sensitivity analysis with deliberate and controlled variations in the model to evaluate responses in the model for plausibility, which can also be called a “structured what-if analysis” (SWI-FRAM) (cf. Hill et al. 2020; MacKinnon et al. 2021). Here, one upstream function will be manipulated to vary its output to understand its impact on the system as well as how this variability can propagate through the system. In terms of predictive validation, this can be used to check whether the variation in the output of the upstream function actually influences the output of the coupled downstream functions while keeping all other functions constant at the same time. This process must be carried out for all direct upstream–downstream couplings of foreground functions in an FRAM model to fully test its predictive validity. This is exemplified in Fig. 2 and Table 1, which will be described in the following. Function A, highlighted in green, is initially manipulated to test the couplings AB, AC, and AD and to see if these couplings actually lead to a change in the output of functions B, C, and D. In the next steps, this procedure is also carried out for the other upstream–downstream couplings of the remaining

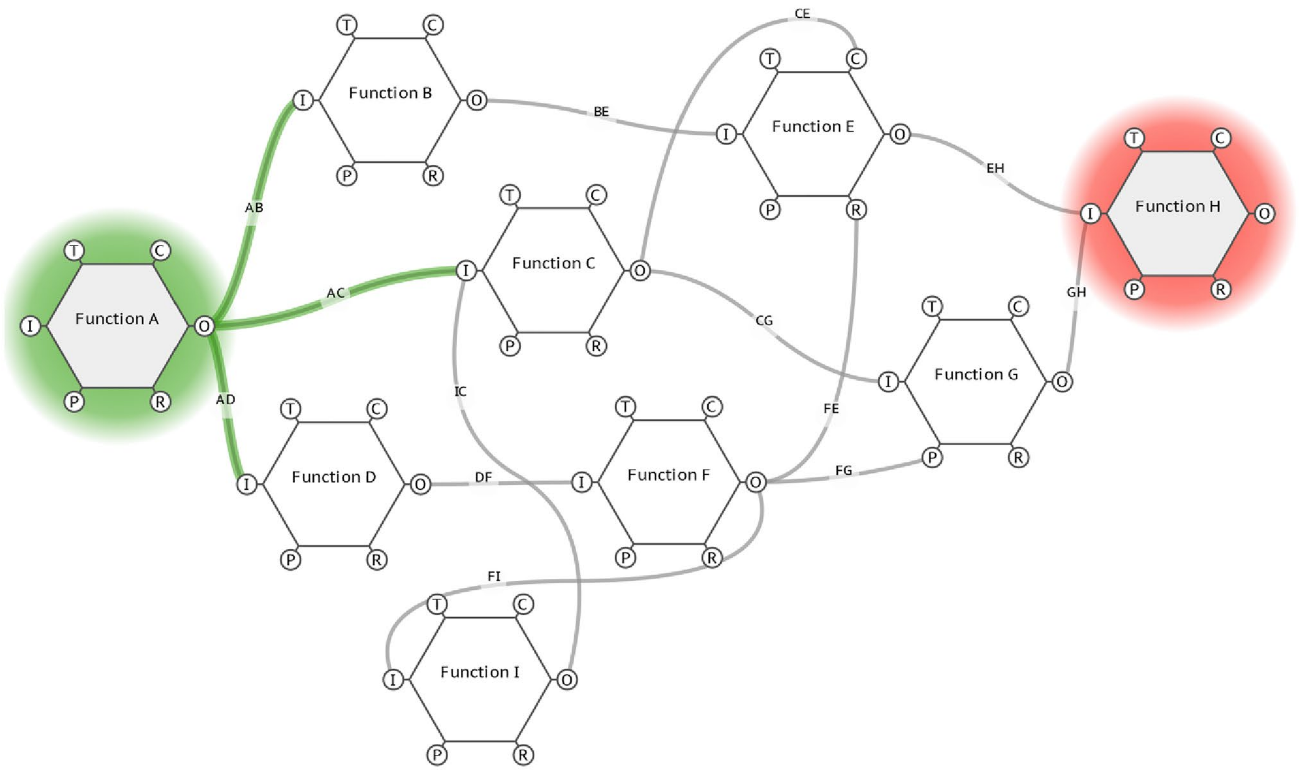


Fig. 2 Fictitious instantiation of an FRAM model with nine functions and thirteen couplings marked through letters. Function A is the start function and function H is the end function, as highlighted in green and red, respectively

Table 1 Assignment of upstream functions, downstream function, and their related couplings with regard to the fictive FRAM model in Fig. 2

Upstream function	Couplings	Downstream functions
A->	AB	-> B
	AC	-> C
	AD	-> D
B->	BE	-> E
C->	CE	->E
	CG	-> G
D->	DF	-> F
E->	EH	-> H
F->	FE	-> E
	FG	-> G
	FI	-> I
G->	GH	-> H
I->	IC	-> C

functions (see Table 1) up to the end function H, highlighted in red. The proof of all direct couplings is sufficient as this automatically explains the indirect effects too, for example, if function A has a direct impact on function D and D in turn on function F, then A also has an indirect effect on F. If the expected effect for one coupling can actually be confirmed, it is valid and if not, the coupling is not relevant and, therefore, invalid. However, this only validates the predictive performance for one instantiation and thus for one specific scenario, which does not mean that the model will be generally valid or invalid for other situations.

The final comparison between expected and actual effect then corresponds to an outcome-based validation, where the predictions of the FRAM model are matched with actual observations in reality. This is where the SDT comes into play, which was pioneered by Stanton and Young (1999a, b) to establish the empirical validity of ergonomics methods as mentioned in Sect. 1.2. This technique divides the method's outputs up into hits (H), misses (M), false alarms (FA), and correct rejections (CR). In the context of FRAM, it provides a method to compare the predicted variability effect of an upstream function to its coupled downstream functions, illustrated through the FRAM model, with the actual observed variability effect in simulator or field tests. In this work, the four events in Fig. 3 are defined as follows:

- *Hits*: predicted variability effect in a downstream function's output through the manipulation of its upstream

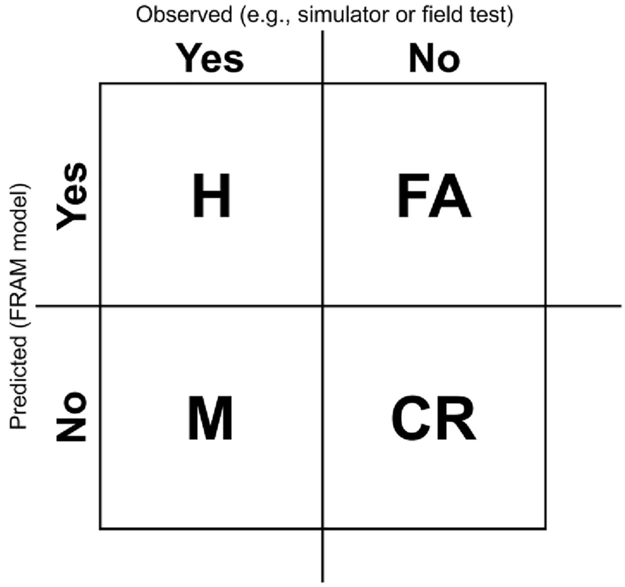


Fig. 3 Signal detection theory (SDT) matrix

function's output by the FRAM model and observed variability effect in a simulator or field test.

- *Misses*: no predicted variability effect in a downstream function's output through the manipulation of its upstream function's output by the FRAM model, but observed variability effect in a simulator or field test.
- *False alarms*: predicted variability effect in a downstream function's output through the manipulation of its upstream function's output by the FRAM model, but no observed variability effect in a simulator or field test.
- *Correct rejections*: no predicted variability effect in a downstream function's output through the manipulation of its upstream function's output by the FRAM model and no observed variability effect in a simulator or field test.

In the following, the four events mentioned above will be explained using examples with the fictive FRAM model in Fig. 2. For instance, we will manipulate the output of function C to test the predictive validity. Potential hits or false alarms could be the couplings CE and CG to its direct downstream functions E and G, with one potential result being that coupling CE is a hit and CG a false alarm. The couplings EH and GH are indirect downstream effects of function C to function H and, therefore, out of the scope as we only measure direct downstream effects. Potential misses or correct rejections could be all the remaining functions that are not indirectly influenced by function C, and where no direct downstream couplings currently exist with function C and thus no variability effects are expected. It has to be proven whether the manipulation of function C has a variability effect on the outputs of the functions A, B, D, F, and I. Potential results could be that the "potential" coupling to function B is a miss and the potential couplings to the functions A, D, F, and I are correct rejections. Several metrics comprising the four events can now be used for the subsequent and concrete evaluation of predictive validity, which will be explained in more detail in Sect. 3.6.

All of the methods described above to demonstrate or increase the reliability, verification, and validity either influence or improve the performance and value of an FRAM model to increase the objective evaluation of research findings by FRAM as depicted in Fig. 1. In the next step, the process of predictive validity will be exemplified through an FRAM model for human and automated driving by Grabbe et al. (2022) to show its credibility as well as the applicability of the previously described predictive validation approach. This is because first, predictive validity represents the highest maxim of validation, and second, reliability, verification, and content validity for the analytical part of the validation have already been implemented by Grabbe et al. (2022) for the model to be examined. Therefore, the evaluative part of the validation is still open and thus addressed in the methods section.

3 Methods

3.1 FRAM model

The FRAM model to be validated in this paper is the FRAM model for overtaking in road traffic created by Grabbe et al. (2022). This model is very large and detailed, comprising 285 functions and including 210 foreground functions, all of which theoretically have to be analysed individually to test the predictive validity of the entire model. This is practically impossible and would go beyond the scope of this work. We, therefore, selected the two functions '*driving free*' (lead vehicle, LV) and '*driving free*' (oncoming vehicle, OV) to demonstrate the predictive validity. Both functions have a major impact on the system or rather the model and basically represent the longitudinal and lateral driving behaviour of LV and OV. The two functions and their couplings as well as their context will be described in more detail in Sects. 3.5 and 3.7.1, and 3.4.2, respectively.

3.2 Research questions

The analysis of the predictive validity of the FRAM model by Grabbe et al. (2022) pursues three research questions:

1) Is the model predictively valid for the basic scenario?

2) Is the model predictively valid for changing environmental conditions?
3) Is the model predictively valid for changing human factors conditions?

3.3 Sample

Forty German participants with valid driving licences took part in this experiment. This sample was divided into two subgroups with twenty participants each for the between-subjects factor levels of time pressure or no time pressure. The mean (*M*) age of the time pressure group was 29.4 years (SD = 14.5 years) with a range from 19 to 75 years, and that of the no time pressure group was 31 years (SD = 14.2 years) ranging from 21 to 72 years. The time pressure group consisted of twelve (60%) men and eight (40%) women, while the no time pressure group consisted of eleven (55%) men and nine (45%) women. In addition, Table 2 gives an overview of a comparison between the no time pressure and time pressure group as regards driving experience and driving style. Based on this data, the two samples can be considered as comparable.

3.4 Apparatus

3.4.1 Driving simulator

The experiment was carried out in the static driving simulator of the Chair of Ergonomics at the Technical University of Munich (see Fig. 4). The simulator consisted of a BMW E64 vehicle mock-up. A high-quality, 6-channel projection system provided a realistic driving environment. Three projectors were used for the front and back view each. The front field of view is approx. 180°. The back view through the mirrors is realised through three separate canvases. SILAB 6.5 of the Würzburg Institute for Traffic Sciences GmbH, with a refresh rate of 60 Hz, was used as the driving simulation software. An additional sound system provided vehicle and environmental sounds.

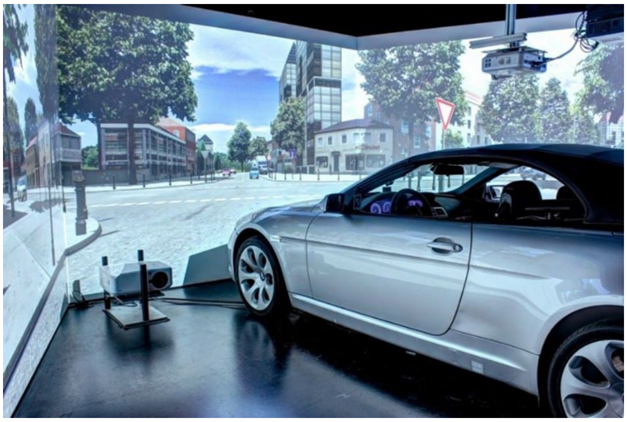


Fig. 4 Static driving simulator

3.4.2 Scenario and experimental track

The scenario of the analysed FRAM model was an overtaking manoeuvre on a rural road as detailed in Grabbe et al. (2022). An ego vehicle (EV) driven by the participant wants to overtake an LV travelling at a speed of 80 km/h on a straight rural road for a distance of 2500 m with no vertical elevation. The maximum speed limit is 100 km/h, overtaking is permitted and no obstructions exist. A rear vehicle (RV) is following the EV, and a line of cars are approaching on the oncoming lane at 100 km/h with different fixed time gaps. There were a total of ten gaps on the straight, with the first four time gaps being 10 s and for the last six gaps 12 s, corresponding to critical and uncritical time gaps according to the mean of 11.5 s (Crawford 1963) and median of 9.9 s (Tapio 2003) found in literature regarding accepted gaps when overtaking passenger cars. The road is 6 m wide, with one lane in each direction and a dotted line in the middle. The road is well constructed and all necessary road markings are in place. There is light vegetation on the side of the road. The weather conditions are sunny and dry. All simulation-controlled vehicles, which are passenger cars, always keep the necessary safety distance to their vehicle in front and comply with the traffic regulations. Before the actual test scenario,

Table 2 Comparison between the no time pressure and time pressure group regarding driving experience and driving style

Measurement	No time pressure group	Time pressure group
Participation in driving simulator studies	*M* = 7.7 (SD = 8.5)	*M* = 10.3 (SD = 24.3)
Mileage [km/year]	*M* = 12,272 (SD = 5,054)	*M* = 12,777 (SD = 5,995)
Driving regularity [daily, weekly, monthly, annually]	Daily 40% Weekly 45% Monthly 15%	Daily 50% Weekly 40% Monthly 10%
Driving style *[5-Likert scale: from (1) very safe to very risky (5)]*	*M* = 2.5 (SD = 0.8)	*M* = 2.5 (SD = 0.9)
Driving pace *[5-Likert scale: from (1) very leisurely to very rapid (5)]*	*M* = 3.3 (SD = 0.6)	*M* = 3.4 (SD = 0.9)
Driving capability *[5-Likert scale: from (1) very inexperienced to very experienced (5)]*	*M* = 4.0 (SD = 0.8)	*M* = 4.1 (SD = 0.9)

the overtaking manoeuvre on the straight rural road, each of the test subjects drove a small winding course for a distance of 2,000 m through a wooded area so that the entire scenario would appear as natural as possible. To get a better overview, the scenario can be divided into five temporal and spatial stages from the EV's point of view (see Fig. 5): following a vehicle in front, swerving into the oncoming lane, passing the leading vehicle, merging back into the starting lane, and getting in the lane again.

3.5 Experimental design

We used a 2×3×3 mixed factorial design for this experiment. The human factors condition (no time pressure or time pressure) was the between-subject factor, while the environmental condition (basic, truck, fog and rain) and the function manipulation (no manipulation, manipulation of driving free LV, manipulation of driving free OV) were within-subject factors (see Fig. 6). Half of the participants experienced time pressure as realised by an expiring time counter in the head-up display, forcing them to overtake as early as possible. The timer was set to expire as soon as the fourth gap had passed, forcing the test persons to overtake in the gaps with the smaller and critical time gaps described in Sect. 3.4.2. The reason for this is that impatient drivers under time pressure tend to reduce the accepted gaps during passing manoeuvres (Pollatschek and Polus 2005). Each test subject drove all nine scenarios, comprising the three different environmental conditions as well as function manipulations, where the scenarios were permuted to mitigate potential sequence and learning effects. The basic condition corresponded to the standard scope of the examined FRAM model, whereas the LV, which was basically a passenger car, was substituted through a truck in the truck condition, and the weather conditions, that were basically sunny and dry, were changed to fog and rain in the third condition. The first three scenarios,

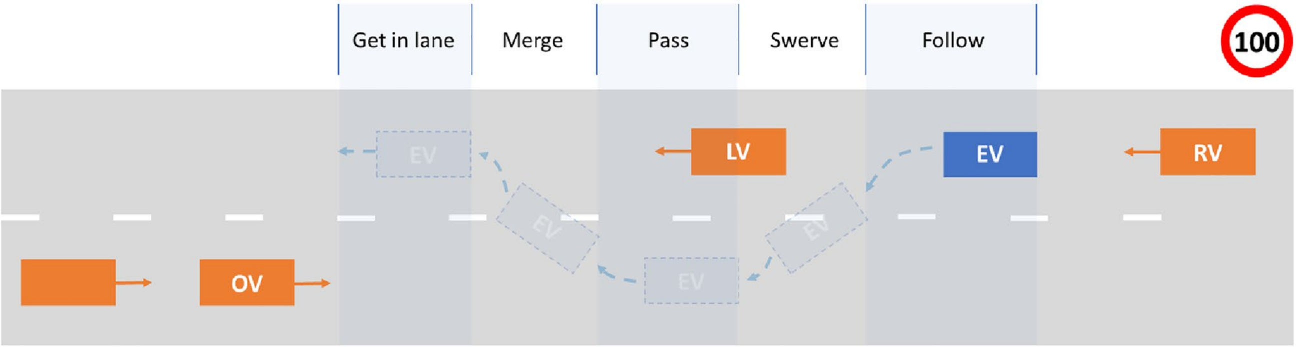
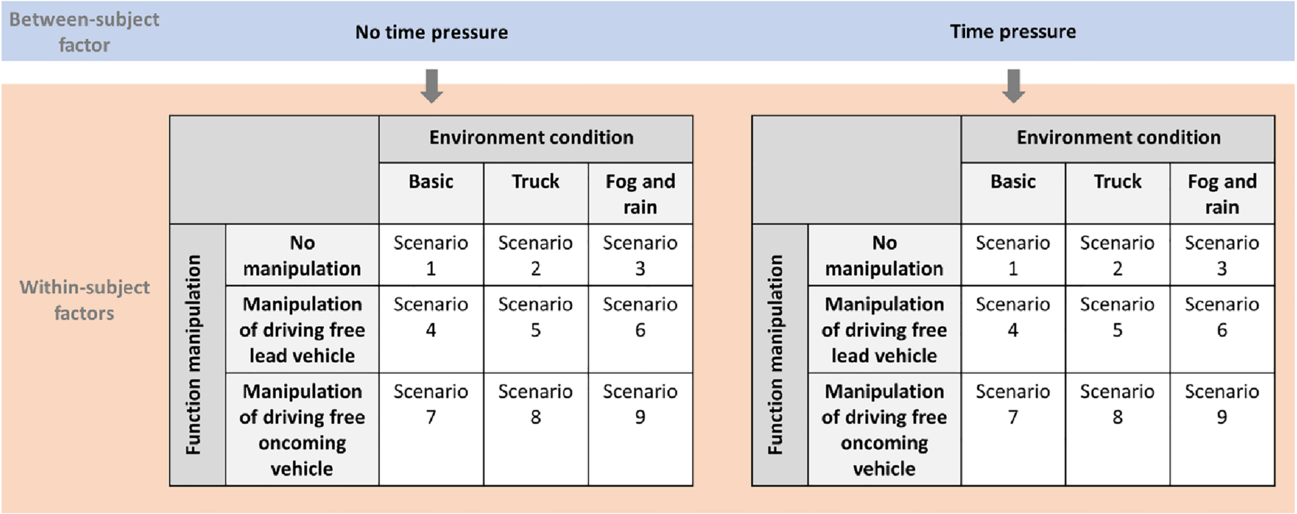


Fig. 5 Schematic illustration of the overtaking scenario comprising different road users/agents and divided into five temporal and spatial stages. *EV* ego vehicle, *LV* lead vehicle, *RV* rear vehicle, *OV* oncoming vehicle, according to Grabbe et al. (2022)



Between-subject factor

Within-subject factors

No time pressure

		Environment condition		
		Basic	Truck	Fog and rain
Function manipulation	No manipulation	Scenario 1	Scenario 2	Scenario 3
	Manipulation of driving free lead vehicle	Scenario 4	Scenario 5	Scenario 6
	Manipulation of driving free oncoming vehicle	Scenario 7	Scenario 8	Scenario 9

Time pressure

		Environment condition		
		Basic	Truck	Fog and rain
Function manipulation	No manipulation	Scenario 1	Scenario 2	Scenario 3
	Manipulation of driving free lead vehicle	Scenario 4	Scenario 5	Scenario 6
	Manipulation of driving free oncoming vehicle	Scenario 7	Scenario 8	Scenario 9

Fig. 6 Illustration of the mixed factorial design

in which no manipulation was implemented, served as references for the three environmental conditions to analyse the predictive validity for the function manipulation of driving free for both the LV and the OV. The manipulation of driving free was realised for the LV by multiple abrupt braking and acceleration as well as repeated lateral offsets, such as "weaving around ", and for the OV by increasing the speed from 100 to 120 km/h, which reduced the time gaps of the first four gaps to 8.33 s and for the last six gaps to 9.99 s, resulting in even more critical time gaps. Finally, scenarios 4 and 7 must be compared with scenario 1, scenarios 5 and 8 with scenario 2, and scenarios 6 and 9 with scenario 3 (see Fig. 6).

3.6 Procedure

Participants were welcomed and informed about the study goals and the procedure. After risks such as nausea and the option of withdrawing from the study without needing to cite reasons were outlined, written consent was obtained. Participants filled out a demographic questionnaire, which also asked for details of their driving experience and driving style. They then drove in the driving simulator for about 10 min to familiarise themselves with the steering, braking, and the driving simulator system. Afterwards, the participants drove a modified basic scenario with no oncoming vehicles and just the LV, with the goal of overtaking this. They were then asked to fill out a questionnaire to rate the timing and precision variability performance of some subjective functions on a 7-Likert scale, as will be more detailed in Sect. 3.7.1. This initial trial run served as a familiarisation for the participants with the basic procedure of the subsequent nine test drives as well as the non-trivial subjective rating of the functions. The actual nine test runs then began, each followed by completing the questionnaire on the subjective functions. Finally, the participants filled out five follow-up questions to rate the perception of the simulated drive. In general, the test subjects were instructed to overtake the LV before the end of the straight by obeying the traffic regulations but also showing her or his most natural and everyday driving behaviour. No restrictions were given regarding overtaking behaviour to ensure idiosyncratic and diverse driving styles. However, an exception exists for the subjects in the group with time pressure who were intentionally instructed to overtake the LV before the timer expires.

3.7 Measures and analysis

3.7.1 Independent and dependent variables

The overall study consisted of three independent variables comprising the function manipulation, environmental condition, and the human factors condition. Moreover, the dependent variables were the performance variability values of several subjective and objective functions in which the performance variability of their outputs, if driving free LV or OV are manipulated, should either change (expected direct downstream effects) or should not change (no expected direct downstream effects) according to the FRAM model by Grabbe et al. (2022) (see Table 3). It should be emphasised that the variability of the outcome or output from a function was measured and not the variability of the function itself. This work only investigated the expected and unexpected downstream couplings of the two manipulated functions to the functions of the agent EV and not to the other agents to test for predictive validity. In the case of expected direct downstream effects, the corresponding functions were assigned to the H/FA category, and in case of no expected direct downstream effects, the corresponding functions were assigned to the M/CR category, according to the application of SDT to FRAM as described basically at the end of Sect. 2.2. Furthermore, in the case of the M/CR category, these functions do not represent all of the potential functions that have to be tested, but only a selection of functions, as otherwise there would be far too many functions for any practical test. Theoretically, these would be all of the remaining functions of the entire model that are not expected to be directly or indirectly influenced by the manipulated functions.

The performance variability of the subjective functions was based on the rating of the timing and precision variability performance in the questionnaire on a 7-Likert scale. Here, the timing was coded as 1 for too early, 3 for on time, 5 for too late, and 7 for not at all, whereas precision was coded as 1 for precise, 4 for acceptable, and 7 for imprecise. The subjects were asked when (timing) or how (precision) they, for example, estimated the distance to OV until they swerved. Finally, the two values for timing and precision were multiplied into one representative value for the performance variability of the subjective functions. By contrast, the performance variability of the objective functions was based on driving data (e.g., speed, lane deviation, and distance between cars) gathered in the driving simulator. For the sake of simplicity, we gathered the performance variability of the objective functions either in terms of timing or precision but not both. An overview of the measurement definitions of each objective function is given in Table 4.

3.7.2 Statistical analysis

To evaluate the predictive validity, the performance variability had to be reduced into the four events of SDT, namely, hits, false alarms, misses, and correct rejections, by comparing the predictions of the model with the observations in the simulator.

Table 3 Assignment of manipulated functions and analysed functions of EV, and their allocation to the type of rating and SDT event category

Manipulated function	Analysed functions of EV	Type of rating	SDT event category
Driving free LV	Check vehicles in front of LV	Subjective	H/FA
	Check LV is not about to change speed		H/FA
	Gauge future driving actions of LV		H/FA
	Check LV is not indicating or about to turn		H/FA
	Maintain an adequate view of the road ahead		H/FA
	Evaluate reasonableness for overtaking		H/FA
	Assess the situation to enter safely		H/FA
	Judge LV's relative speed to OV		H/FA
	Judge LV's speed		H/FA
	Judge available passing time		H/FA
	Determine pass can be completed		H/FA
	Observe road behind		M/CR
	Check for safe distance to merge		M/CR
	Judge first OV's speed		M/CR
	Judge distance from first OV		M/CR
	Maintain headway separation	Objective	H/FA
	Keep in lane		H/FA
	Position car to the right		H/FA
	Position car to the left		H/FA
	Reduce headway from normal following		H/FA
	Avoid tailgating and intimidating LV		H/FA
	Adjust speed to that of LV		H/FA
	Adopt overtaking position		H/FA
	Swerve completely to the oncoming lane		H/FA
	Accelerate LV decisively		H/FA
	Merge back into starting lane		H/FA
	Merge progressively into starting lane		H/FA
	Comply with the speed limit		M/CR
Driving free OV	Judge first OV's speed	Subjective	H/FA
	Judge LV's relative speed to OV		H/FA
	Judge available passing time		H/FA
	Determine pass can be completed		H/FA
	Assess the situation to enter safely		H/FA
	Judge distance from first OV		M/CR
	Judge LV's speed		M/CR
	Observe road behind		M/CR
	Check for safe distance to merge		M/CR
	Accelerate LV decisively	Objective	H/FA
	Merge back into starting lane		H/FA
	Merge progressively into starting lane		H/FA
	Comply with the speed limit		M/CR
	Maintain headway separation		M/CR
	Keep in lane		M/CR

First, the mean and standard deviation of the performance values were calculated for the scenarios 1–3 (as these form the respective reference for testing for differences in performance variability as described in Sect. 3.5) for each analysed function per between-subject factor group, from which the 95% confidence interval was calculated to define a "normal" everyday variability range. In medicine, one also speaks of normal ranges, which are defined for blood pressure or blood

Table 4 Overview of the measurement definitions of each objective function

Objective function	Stage	Phenotype	Definition
Maintain headway separation	Follow	Precision	The average distance between EV and LV in the period, where the straight begins and the driver of EV starts to swerve, indicated by the left activated indicator or the steering angle
Keep in lane	Follow	Precision	The average absolute lane deviation between of EV in the period, where the straight begins and the driver of EV starts to swerve
Position car to right/left	Follow	Precision	The average gap to the left/right lane edge of in the period, where the straight begins and the driver of EV starts to swerve
Reduce headway from normal following	Swerve	Precision	The average distance between EV and LV in the period, where the driver of EV starts to swerve and driving completely in the oncoming lane, indicated by the left activated indicator or the steering angle, and the lane index showing in which lane EV is driving, respectively
Avoid tailgating and intimidating LV	Swerve	Precision	The distance between EV and LV at the last point, where the driver of EV is driving in the starting lane and already has started to swerve
Adjust speed to that of LV	Swerve	Precision	The average speed difference between EV and LV in the period, where the straight begins and the driver of EV starts to swerve
Adopt overtaking position	Swerve	Precision	The sum of the speed of EV, absolute lane deviation of EV, and the distance between EV and LV at the point, where the driver of EV starts to swerve
Swerve completely to oncoming lane	Swerve	Timing	The time difference between starting to swerve and driving completely in the oncoming lane
Accelerate LV decisively	Pass	Precision	The average speed of EV in the period, where the driver of EV starts to drive completely in the oncoming lane and starts to merge, indicated by the lane index showing in which lane EV is driving, and the right activated indicator or the steering angle, respectively
Merge back into starting lane	Pass	Precision	The number of times the driver of EV merged back into the starting lane even though the driver has already swerved into the oncoming lane to overtake
Merge progressively into starting lane	Merge	Timing	The time difference between starting to merge and driving completely in the starting lane
Comply with the speed limit	All	Precision	The average speed difference between EV's speed and the speed limit in the period, where the straight begins and the driver of EV is driving completely in the starting lane again after passing LV

sugar, for example, to distinguish healthy patients from sick patients. Afterwards, the difference between the upper/lower limit of the confidence interval and the mean was calculated, which reflects a maximum positive or negative everyday fluctuation in performance that is normal and thus should not be regarded as a significant performance variability.

We then calculated the absolute differences between the intraindividual performance values of scenario 4 and 7 to 1, scenario 5 and 8 to 2, and scenario 6 and 9 to 3 for each analysed function as we were interested in both the positive and negative direction of the performance variability. In the next step, one-sided one-sample *t* tests with a p-value of 5% were used to determine whether the sample mean of the absolute differences in performance of, for example, scenario 4 to 1 was statistically greater than the respective maximum value of everyday fluctuation in performance. The Wilcoxon signed-rank test was used as an alternative when the statistical requirements for the one-sample t-test were not met. If the p-value was lower than 5%, then a significant performance variability in the analysed function in the respective scenario was assumed, otherwise not.

From this, it was possible to finally assess which of the four events according to SDT applies per analysed function, group and scenario. Subsequently, the number of the four events per manipulated function (driving free LV and OV), human factors condition (time pressure, no time pressure) and environmental condition (basic, truck, rain and fog) were calculated. Based on this, the accuracy, H-rate (HR), and CR-rate (CRR) were calculated to be able to prove the predictive validity. We decided to use the accuracy and not the Matthews (1975) correlation coefficient (MCC), which is generally recommended by Stanton and Young (1999a, 2003) and successfully applied, for example, by Stanton et al. (2021a; b, c, d) and Hulme et al. (2021a; b, c), as an appropriate statistical metric to validate human factors methods in binary classification problems. The reasons are twofold. First, the analysed FRAM model is clearly complex with a wide scope, and according to Stanton and Young (2003), the wider the scope of the method or model, the more difficult it is to obtain favourable data on validity performance, so it would be detrimental to use a harsh metric like the MCC. Second, the true positive results should be favoured over the true negative results as considerably more

H than CR can be identified in an FRAM model validation due to the practical limitations mentioned in Sect. 3.7.1. The accuracy score tends to favour positive cases (Baber and Young 2022). With this in mind, accuracy seemed to be more appropriate than MCC to obtain a high validity score, because it is quite difficult to obtain a high score through good prediction results in only all four of the confusion matrix categories. Nevertheless, as using the accuracy alone as a single value to prove predictive validity could be misleading in the case of imbalanced classification data sets (cf. Chicco and Jurman 2020), we also considered the HR, and especially CRR, to achieve a broader and more detailed analysis.

The numerical value of accuracy represents the proportion of true or expected results (both true positive (H) and true negative (CR)) and was calculated as follows (1):

$$\text{Accuracy} = \frac{\text{H} + \text{CR}}{\text{H} + \text{FA} + \text{M} + \text{CR}} \tag{1}$$

HR or sensitivity represents the proportion of true positives or expected and observed results and was calculated as follows (2):

$$\text{HR} = \frac{\text{H}}{\text{H} + \text{M}} \tag{2}$$

CRR or specificity represents the proportion of true negatives or not expected and not observed results and was calculated as follows (3):

$$\text{CRR} = \frac{\text{CR}}{\text{FA} + \text{CR}} \tag{3}$$

All three metrics are expressed along a percentage scale ranging from 0 to 100. Ultimately, a criterion for acceptable levels of predictive validity has to be considered, as there is no universally accepted measure. A review of reliability and validity levels found that, across 25 studies, the average value used to indicate acceptable percentage agreement was 76%, with a range of 70–88% (Olsen 2013). As described in Sect. 2.1, validation is gradual rather than binary. Thus, a single value indicating that an FRAM model is predictively valid or not seems to be inappropriate. Rather, a more differentiated approach was used in this work, defining different levels for predictive validity from *poor* to *almost perfect* according to the reliability result levels applied to SDT by Olsen (2013) (see Table 5). However, to answer the research questions in Sect. 3.2, we additionally defined a value for sufficient predictive validity, which was set at 70%. We have chosen this value, because it defines first, the minimum of acceptable percentage agreement (Olsen 2013), and second, the median of the category of substantial predictive validity, which should be the minimum category to aim for (see Table 5).

Table 5 Levels for predictive validity associated with percentages of selected metrics according to Olsen (2013)

Predictive validity level	Percentage of accuracy, HR, and CRR
Poor	0
Slight	> 0–20
Fair	21–40
Moderate	41–60
Substantial	61–80
Almost perfect	81–100

4 Results

This section presents the results according to the three different research questions defined in Sect. 3.2. An overview of the results of the SDT event category for every analysed function per manipulated function, with a differentiation between human factors and environmental conditions, is shown in Table 6.

4.1 Predictive validity for the basic scenario

Figure 7 shows the comparison of the accuracy, HR and CRR associated with the predictive validity levels (see Table 5) between the environmental and human factor conditions with the manipulated function of driving free LV. Furthermore, the 70% threshold as the value for sufficient predictive validity is indicated by a horizontal dashed red line. For the basic scenario in the no time pressure group, the accuracy, HR, and CRR account for 79%, 81% and 0%, respectively. The accuracy and HR lie above the sufficient predictive validity, reaching a substantial and almost perfect predictive validity level, respectively. However, the predictive validity level of the CRR is poor. In total, there are six (21%) functions that do not meet expectations: '*observe road behind*', '*check for safe distance to merge*', '*judge first OV's speed*', '*judge distance from first OV*' (all are M instead of CR) as subjective functions and '*merge back into starting lane*' (FA instead of H) and '*comply with the speed limit*' as objective functions (M instead of CR). It is noticeable that the false predictions are mainly based on misses.

Figure 8 is the same as Fig. 7, but for the manipulated function of driving free OV. Here, the accuracy, HR, and CRR account for 53%, 56% and 50%, respectively, for the basic scenario in the no time pressure group. Therefore, all three metrics lie below the sufficient predictive validity and achieve a moderate predictive validity level. In total, there are seven (47%) functions that do not meet expectations: '*judge LV's relative speed to OV*' (FA instead of H), '*judge distance from first OV*', '*judge LV's speed*', and '*observe road behind*' (all are M instead of CR) as subjective

Table 6 Assignment of manipulated functions and analysed functions of EV, and their respective results of SDT event category, with a differentiation between human factors and environmental conditions

Manipulated function	Analysed functions of EV	No time pressure			Time pressure		
		Basic	Truck	Rain/fog	Basic	Truck	Rain/fog
Driving free LV	Check vehicles in front of LV	H	H	FA	H	H	H
	Check LV is not about to change speed	H	H	H	H	H	H
	Gauge future driving actions of LV	H	H	H	H	H	H
	Check LV is not indicating or about to turn	H	H	H	H	H	H
	Maintain an adequate view of the road ahead	H	H	H	H	H	H
	Evaluate reasonableness for overtaking	H	FA	H	H	H	H
	Assess the situation to enter safely	H	H	H	FA	FA	FA
	Judge LV's relative speed to OV	H	H	H	H	H	H
	Judge LV's speed	H	H	H	H	H	H
	Judge available passing time	H	H	H	H	H	H
	Determine pass can be completed	H	H	H	H	H	H
	Observe road behind	M	CR	M	M	M	M
	Check for safe distance to merge	M	CR	M	M	M	CR
	Judge first OV's speed	M	M	M	M	M	M
	Judge distance from first OV	M	M	M	CR	CR	CR
	Maintain headway separation	H	H	FA	H	H	FA
	Keep in lane	H	H	H	H	H	FA
	Position car to the right	H	FA	FA	H	H	H
	Position car to the left	H	FA	FA	H	H	H
	Reduce headway from normal following	H	H	H	H	H	H
	Avoid tailgating and intimidating LV	H	H	H	H	H	H
	Adjust speed to that of LV	H	H	H	H	H	FA
	Adopt overtaking position	H	H	H	H	H	H
	Swerve completely to the oncoming lane	H	H	H	H	H	FA
	Accelerate LV decisively	H	H	H	H	H	H
	Merge back into starting lane	FA	FA	FA	FA	FA	FA
	Merge progressively into starting lane	H	H	H	H	H	H
	Comply with the speed limit	M	M	M	M	M	M
Driving free OV	Judge first OV's speed	H	FA	H	H	H	FA
	Judge LV's relative speed to OV	FA	H	H	H	H	FA
	Judge available passing time	H	FA	H	H	H	H
	Determine pass can be completed	H	H	H	H	H	H
	Assess the situation to enter safely	H	FA	H	FA	H	FA
	Judge distance from first OV	M	M	M	CR	CR	CR
	Judge LV's speed	M	M	M	CR	CR	M
	Observe road behind	M	M	M	M	M	CR
	Check for safe distance to merge	CR	M	M	M	M	M
	Accelerate LV decisively	FA	FA	H	FA	H	FA
	Merge back into starting lane	FA	FA	FA	FA	FA	FA
	Merge progressively into starting lane	H	FA	H	FA	H	H
	Comply with the speed limit	CR	M	M	M	CR	CR
	Maintain headway separation	CR	CR	CR	CR	M	CR
	Keep in lane	M	CR	M	M	M	CR

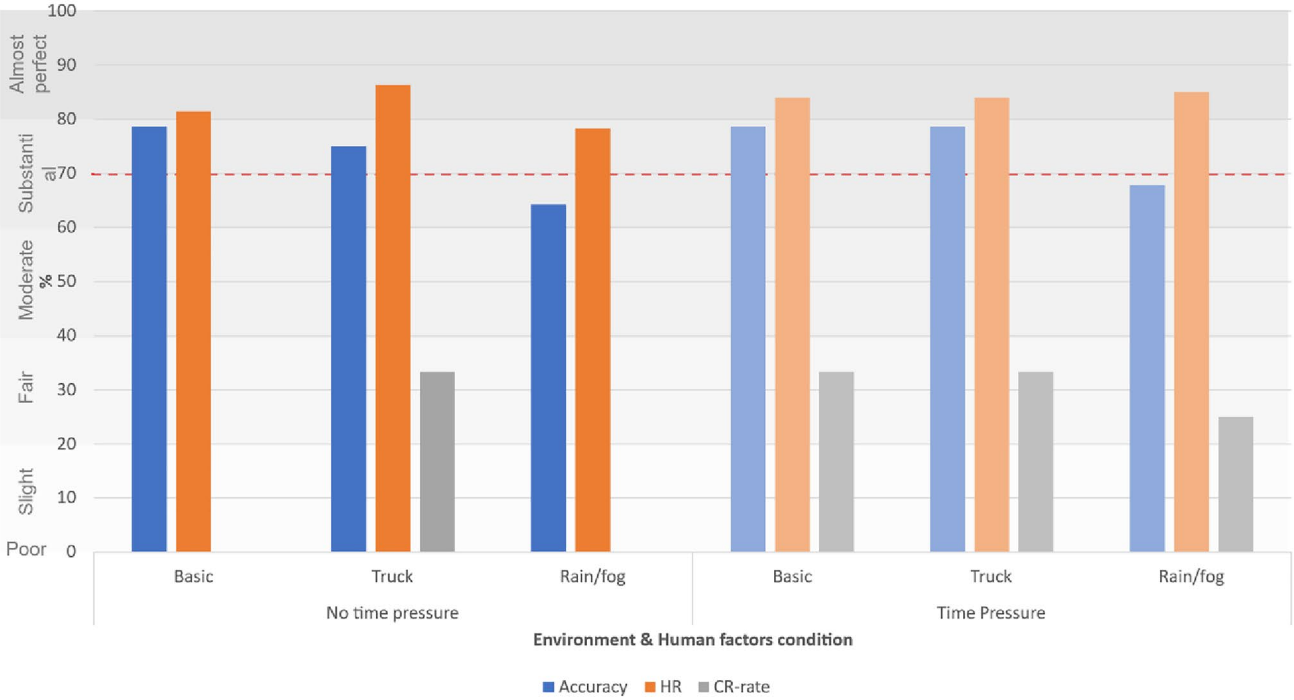


Fig. 7 Comparison of the accuracy, HR, and CRR associated with the predictive validity levels between the environmental and human factors conditions for the manipulated function of driving free LV

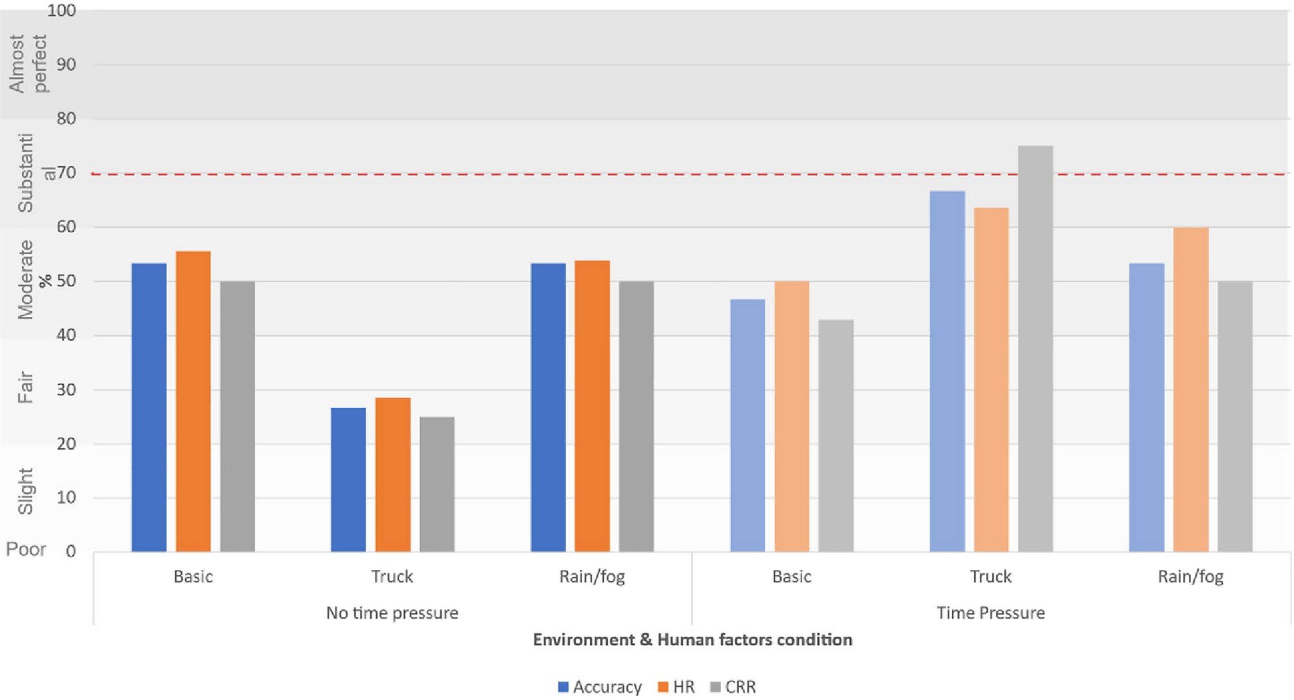


Fig. 8 Comparison of the accuracy, HR, and CRR associated with the predictive validity levels between the environmental and human factors conditions for the manipulated function of driving free OV

functions and ‘*accelerate LV decisively*’, ‘*merge back into starting lane*’ (both are FA instead of H) and ‘*keep in lane*’ (M instead of CR) as objective functions. There is a roughly equal distribution of false alarms and misses here.

Besides, a comparison of the accuracy, HR, and CRR between objective and subjective functions shows no clear differences in terms of the type of rating (see Fig. 9). In most cases, the differences amount to a maximum of 10% and alternate, so that sometimes the objective functions achieve a higher value than the subjective functions and vice versa.

4.2 Predictive validity for other environmental conditions

The accuracy, HR, and CRR account for 75%, 86% and 33%, respectively, for the truck scenario in the no time pressure group with the manipulated function of driving free LV (see Fig. 7). Thus, the accuracy and HR lie above the sufficient predictive validity, reaching a substantial and almost perfect predictive validity level, respectively. However, the predictive validity level of the CRR is fair. These results are similar to the ones of the basic scenario. For the rain/fog scenario in the no time pressure group with the manipulated function of driving free LV, the accuracy, HR, and CRR account for 64%, 78% and 0%, respectively (see Fig. 7). Therefore, the accuracy lies below and the HR lies above the sufficient predictive validity, both reaching a substantial predictive validity level, respectively. However, the predictive validity level of the CRR is poor. Slight differences can thus be determined compared to the basic scenario.

For the truck scenario in the no time pressure group with the manipulated function of driving free OV, the accuracy, HR, and CRR account for 27%, 29% and 25%, respectively (see Fig. 8). This means that all three metrics lie below the sufficient predictive validity, reaching a fair predictive validity level. Compared to the basic scenario, this is one level lower. In contrast, the accuracy, HR, and CRR account for 53%, 54% and 50%, respectively, for the rain/fog scenario in the no time pressure group with the manipulated function of driving free OV (see Fig. 8). Thus, all three metrics lie below the sufficient predictive validity, reaching a moderate predictive validity level. These results are similar to those for the basic scenario.

If we consider the functional level of each analysed function and respective changes to the SDT event category between the environmental conditions in relation to the basic scenario for the no time pressure group in Fig. 10, we see that in the truck scenario, 18% of the analysed functions for the manipulated function of driving free LV and 53% of the analysed functions for the manipulated function of driving free OV deviate relative to the SDT event category. In the rain/fog scenario, 14% of the analysed functions for the manipulated function of driving free LV and 33% of the analysed functions for the manipulated function of driving

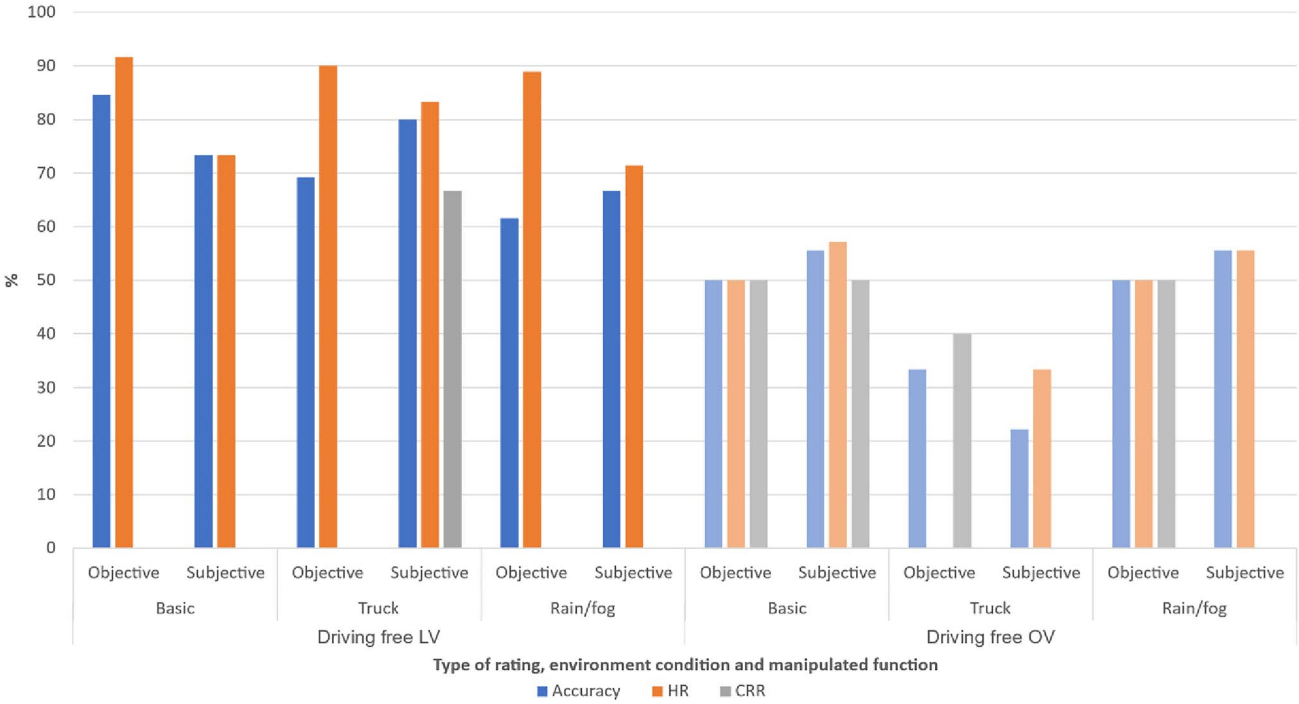


Fig. 9 Comparison of the accuracy, HR, and CRR between the subjective and objective type of rating, with a differentiation between the environmental conditions and manipulated function

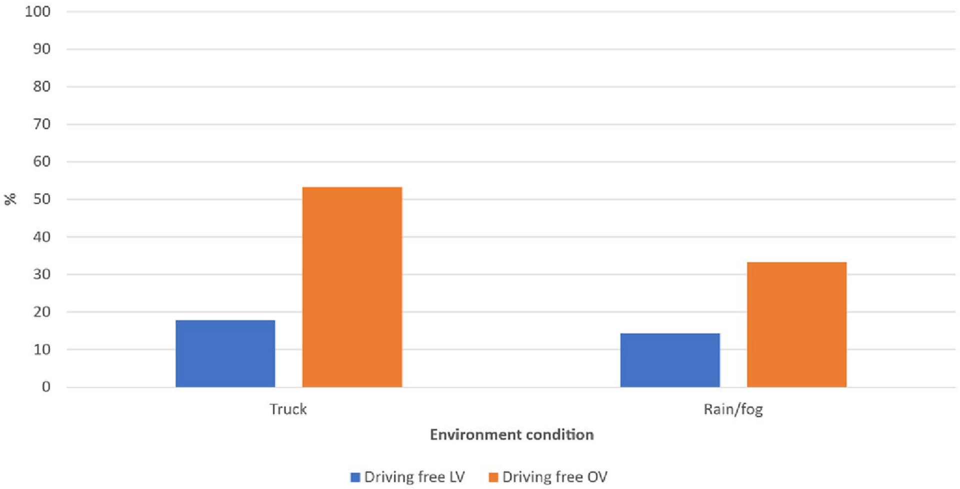


Fig. 10 Relative frequencies of deviations in the SDT event categories within the analysed functions between the manipulated function of driving free LV and OV for the no time pressure group in the truck, and fog and rain scenario, each compared to the basic scenario

free OV deviate relative to the SDT event category. Hence, we can see far greater differences in the predictive validity on the functional level when the environmental condition changes compared to the results of the three metrics shown above, especially with the manipulated function of driving free OV. The number of deviations between the two environmental conditions are similar in the case of the manipulation of driving free LV and different in the case of the manipulation of driving free OV, where the truck scenario shows considerably more deviations than the rain/fog scenario.

4.3 Predictive validity for other human factors conditions

First, we present the results for the manipulated function of driving free LV. For the basic scenario in the time pressure group, the accuracy, HR, and CRR account for 79%, 84% and 33%, respectively (see Fig. 7). Thus, the accuracy and HR lie above the sufficient predictive validity, reaching a substantial and almost perfect predictive validity level, respectively. However, the predictive validity level of the CRR is fair. These results are similar to those for the basic scenario in the no time pressure group, except for the CRR, which is two levels higher. The accuracy, HR, and CRR account for 79%, 84% and 33%, respectively, for the truck scenario in the time pressure group (see Fig. 7). Therefore, the accuracy and HR lie above the sufficient predictive validity, reaching a substantial and almost perfect predictive validity level, respectively. However, the predictive validity level of the CRR is fair. These results are similar to those for the truck scenario in the no time pressure group. The accuracy, HR, and CRR account for 68%, 85% and 25%, respectively, for the rain/fog scenario in the time pressure group (see Fig. 7). Hence, only the HR lies above the sufficient predictive validity, reaching an almost perfect predictive validity level. However, the predictive validity levels of the accuracy and CRR are substantial and fair, respectively. These results are similar to those for the rain/fog scenario in the no time pressure group, except for the CRR, which is two levels higher.

The results for the manipulated function of driving free OV are presented below. The accuracy, HR, and CRR account for 47%, 50% and 43%, respectively, for the basic scenario in the time pressure group (see Fig. 8). This means that all three metrics lie below the sufficient predictive validity level, reaching a moderate predictive validity level. These results are similar to those for the basic scenario in the no time pressure group. The accuracy, HR, and CRR account for 67%, 64% and 75%, respectively, for the truck scenario in the time pressure group (see Fig. 8). Therefore, all three metrics reach a substantial predictive validity level, but only the CRR achieves the sufficient predictive validity threshold. These results differ from those for the truck scenario in the no time pressure group, since all three metrics are one predictive validity level higher. The accuracy, HR, and CRR account for 53%, 60% and 50%, respectively, for the rain/fog scenario in the time pressure group (see Fig. 8). Thus, all three metrics lie below the sufficient predictive validity, reaching a moderate predictive validity level. These results are similar to those for the rain/fog scenario in the no time pressure group.

On the functional level of each analysed function and their possible respective changes to the SDT event category between the human factors conditions relative to each environmental condition in Fig. 11, we can see a trend of increasing deviations for the manipulated function of driving free LV, starting from the basic scenario (7%), via the truck scenario (25%) to the rain/fog scenario (32%). In contrast, the deviations for the basic scenario (47%), the truck scenario (67%) and the rain/fog scenario (53%) are similar in the basic and rain/fog scenario, whereas the truck scenario has a clearly greater deviation in the case of the manipulation of

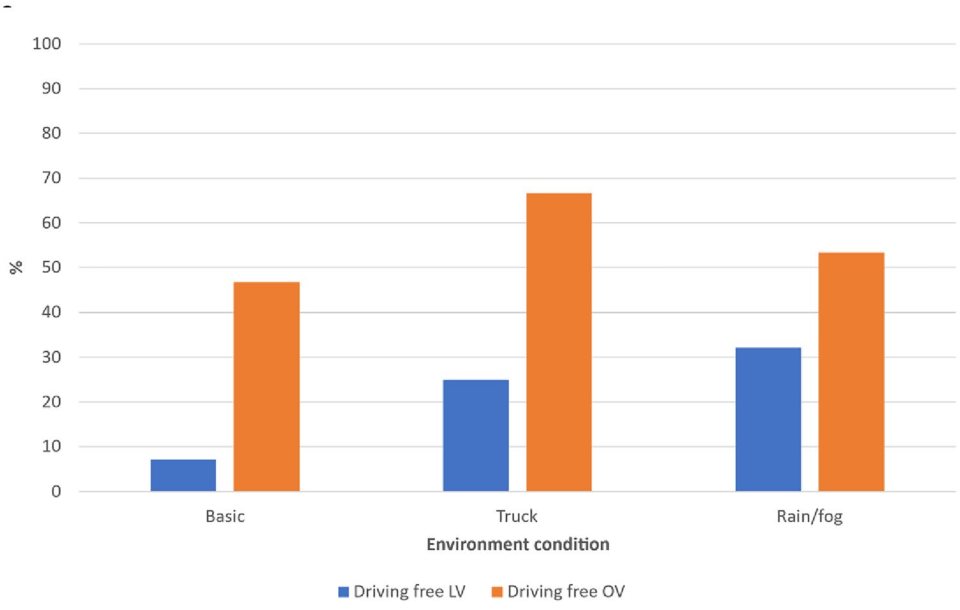


Fig. 11 Relative frequencies of deviations in the SDT event categories within the analysed functions between the manipulated function of driving free LV and OV, comparing each environmental condition between the no time/time pressure groups

driving free OV. Moreover, the number of deviations is considerably higher than for driving free LV. We can see much greater differences in predictive validity on the functional level when the human factors condition changes compared to the results of the three metrics shown above, the same as with the environmental conditions. However, the number of deviations is below 10% in the case of the manipulation of driving free LV for the basic scenario, which should be acceptable, whereas the remaining cases represent clearly higher deviations.

5 Discussion

The aim of this paper is first, to define a more formal approach to achieving and demonstrating the reliability and validity of an FRAM model, and second, to apply this formal approach partly to an existing FRAM model so as to prove its validity and to evaluate the applicability of this approach. In the first part of the paper, a formal approach was derived by transferring both the general understanding and definitions of reliability and validity along with concrete methods and techniques that have been applied in other research areas, or specifically to HFE methods, to the concept of FRAM. In the second part, the predictive validity, which is one part of the formal approach to demonstrate the evaluative part of the validity of an FRAM model, was assessed for a specific FRAM model by Grabbe et al. (2022) in a driving simulator study. Predictive validity represents the highest maxim of validation and the remaining parts of the formal approach had already been applied by Grabbe et al. (2022). Finally, the results of the study have to be discussed so as to prove the credibility of the analysed FRAM model, to cover methodological limitations and to evaluate the utility and applicability of the approach in general.

5.1 Predictive validity of the analysed FRAM model

The research questions from Sect. 3.2 have to be answered in the following to assess the predictive validity of the analysed FRAM model. The following rule applies here: if both the accuracy and HR are sufficient, then predictive validity can be assumed as the true positive results are favoured over the true negative results.

The FRAM model is predictively valid for the basic scenario in the case of the manipulation of driving free LV, because the accuracy and HR are sufficient and reach at least a substantial predictive validity level with high sensitivity. However, the CRR is poor due to several misses, indicating a low specificity. In contrast, the FRAM model is not enough predictively valid for the basic scenario in case of the manipulation of driving free OV, because all three evaluation criteria are insufficient and only reach a moderate level of predictive validity. Overall, the results show that the predictive validity of the FRAM model for the basic scenario is limited, in particular in its specificity, indicating deficiencies in the credibility of the examined FRAM model. In total, the couplings to 13 functions have to be updated. The validation performance of the FRAM model is comparable with the better performing HFE methods in terms of validation (Stanton and Young 1999a; Stanton et al. 2013) only in case of the manipulation of driving free LV, except for the low specificity. Some of the best methods in the field, for example, are associated with the prediction of human error (Baber and Stanton 1996; Harris et al. 2005; Stanton et al. 2009).

These better-performing methods typically achieve validity statistics above 0.8 (Stanton et al. 2021b).

When comparing the differences between the environmental conditions, the results show that the predictive validity is comparable between the three different conditions for both manipulation cases, apart from the truck condition for manipulation of driving free OV, though the deviations in the SDT event categories within each analysed function are clearly high. Therefore, the FRAM model is not predictively valid for other environmental conditions. When the human factors condition is changed, the results indicate that the predictive validity is similar for the two conditions with every environmental condition and both manipulation cases, except the truck condition for manipulation of driving free OV. However, the deviations in the SDT event categories within each analysed function are once again clearly high, except the basic condition for manipulation of driving free LV, which shows low deviations. Hence, the FRAM model is predictively valid for other human factors conditions in the case of the basic scenario with the manipulation of driving free LV, but not for the remaining cases. Consequently, it can be said that a generalisation of the predictive validity of an FRAM model is greatly limited so that an FRAM model has to be adapted to changes in both the environmental as well as human factors conditions, especially if conditions are combined. This is not surprising as an FRAM model can only be validated for specific instantiations, and if the conditions change, the instantiation will change and the model will then have to be adapted and no generalisation will be possible. Against this background, it can also be assumed that the effects of shared and traded control (Sheridan 1992) between the driver and an automation system by enhancing the scenario through an interaction of the driver with an advanced driver assistance system (ADAS), e.g., lane-keeping assist (LKA) and adaptive cruise control (ACC), cannot be validly predicted without adapting the FRAM model. Here, the effects and their prediction of conflict or confusion situations between the two agents would be of particular interest. For example, a dangerous situation can occur when the driver performs a lane change without activating the turn signal, which the LKA could then interpret as an unintentional drift and decide to return the car to the main lane. In addition, this could lead to a decrease in trust or an increased stress level which in turn degrade the driving performance or potentially result in a deactivation of the ADAS by the driver. Such conflicting decisions are called human–machine dissonance when contradictory information exists between humans' and autonomous systems' knowledge, from information processing to actions on a controlled process (Vanderhaegen 2021), and these discrepancies can affect human factors and produce, e.g., discomfort, overload, or stress (Vanderhaegen 2014, 2016). In the FRAM model examined, these conflicts are already present in the form of human–human dissonances, e.g., the manipulation of driving free for the LV by multiple abrupt braking and acceleration could be interpreted by the driver of EV in two main aspects: either that LV is reacting to an obstacle or leading vehicle or that the driver of LV is drunk. Here, the result of the interpretation probably leads to two different reactions of the driver of EV which can lead to dangerous situations. For instance, the EV's driver gauges future driving actions of LV which is significantly facilitated when LV is reacting to the traffic in front and the EV's driver has a clear sight compared to the situation when LV's driver is drunk as her/his driving behaviour is random. In addition, the strange driving behaviour of LV may affect human factors by causing anxiety or increased stress for the driver of EV. This could be a possible reason for false expectations in the FRAM model. Thus, various behavioural changes can be triggered in the system, which primarily affects human factors, which in turn cause behavioural adaptations in the system through interdependencies (cf. Wege et al. 2014). As previously described, changing human factors conditions and their effects cannot be fully predicted with the FRAM model. It would, therefore, be relevant in the future to adapt the FRAM model in this direction and to prove whether the FRAM model is valid in the context of interaction between drivers and ADAS. This appears to be especially important given the increasing introduction of such automation systems into the road system and their risk assessment. In principle, possible conflicts in the sense of dissonance can be represented and identified in an FRAM model via the couplings between the functions when analysing them in the form of "what-if analyses" (Hill et al. 2020; MacKinnon et al. 2021) to understand how a potential conflicting coupling affects several downstream functions and how this propagates through the system.

5.2 Limitations

Some methodological limitations are discussed in the following, including the sample, the driving simulator validity as well as the test setup, the statistical analysis, and the theoretical concept of the predictive validation approach.

The participant characteristics play a role in a driving simulator study (Blana 1996). The narrower sample here might not represent the entire driver population, which is why the evaluation of predictive validity based on performance variability is only valid to a limited extent. Nevertheless, the sample size can be considered as sufficient for the narrower population, since a sample size of 20 test drivers, for example, is sufficient to test the controllability of driver assistance systems according to ISO 26262 (2018).

If we take a closer look at the perceptions in the simulator and compare these between the two different groups (see Table 7), we see that the feeling of time pressure cannot be

Table 7 Comparison of the no time pressure and time pressure group with regard to perception in the driving simulator

Measurement	No time pressure group	Time pressure group
Realistic simulation behavior *[5-Likert scale: from (1) very realistic to very unrealistic (5)]*	*M* = 2.5 (SD = 1.0)	*M* = 2.6 (SD = 1.1)
Realistic driving behaviour of other road users *[5-Likert scale: from (1) very realistic to very unrealistic (5)]*	*M* = 2.8 (SD = 1.1)	*M* = 2.6 (SD = 0.9)
Equivalent overtaking manoeuvers in real life *[5-Likert scale: from (1) very equal to very unequal (5)]*	*M* = 3.1 (SD = 1.4)	*M* = 3.0 (SD = 1.2)
The feeling of time pressure *[5-Likert scale: from (1) very strong to very weak (5)]*	*M* = 3.2 (SD = 1.0)	*M* = 3.5 (SD = 1.0)
Efficiency/safety trade-off of overtaking manoeuver *[5-Likert scale: from (1) efficient to safe (5)]*	*M* = 2.3 (SD = 1.0)	*M* = 2.2 (SD = 1.1)

assumed for the time pressure group as the value is even higher than in the no time pressure group. Furthermore, there are no clear differences in the efficiency/safety trade-off between both groups, which is in contrast to the expectation that the no time pressure group should drive as safely as possible, and the pressure group more efficiently. Thus, it is questionable whether the measures to generate time pressure actually worked. According to Rastegary and Landy (1993), time constraints such as those used in this study may be insufficient for eliciting time pressure per se. These authors attested that not having enough time creates a feeling of time pressure only if the time limit is compulsory and if violating the time limit leads to a sanction. Although the time limit was compulsory, it did not lead to any sanctions. Nevertheless, almost all the drivers in the group with time pressure tried to overtake seriously before the time expired and actually overtook. Therefore, it can be argued that the main intention, to simulate impatient drivers under time pressure who tend to reduce the accepted gaps while performing passing manoeuvres, was accomplished.

According to Grabbe et al. (2022), a driving simulator is an appropriate tool for assessing performance variability in terms of action functions at the operational level, but not for perception and cognitive functions, where we have chosen a mix of objective and subjective measurement of performance variability. This leaves room for criticism, as the variables selected to measure performance and the data collection measures affect the driving simulator validity (Blana 1996; Kaptein et al. 1996). In particular, the variability measured subjectively could be limited in representing the real performance variability as the self-awareness of humans about their performance may be biased. However, this does not appear justified, since no great differences could be found between the type of rating and level of validity. Another issue is the definition of performance variability for the objective functions. For the sake of simplicity, their variability was based either on a timing or a precision metric, but not both. Furthermore, the variability measurement of each objective function was subjectively defined. Thus, it is uncertain whether the variability that is measured objectively completely fits the real performance of a respective function.

The driving simulator could, on the whole, have a great impact on the validity results as the validity of driving simulators is an ongoing concern. Typically, they are valuable tools in road safety and human factors research and have been used to assess a variety of driving performances (Mullen et al. 2011) by providing a safe and controllable environment to investigate driver behaviour ethically, effectively, and efficiently (Larue et al. 2018). However, simulators will never reproduce reality accurately and tend to compromise real-life situations (Espié et al. 2005). For instance, participants will probably not drive normally, because they perceive the driving task as a game, experience motion sickness, or find the driving task unrealistic (Larue et al. 2018). In particular, simulator validity depends on the simulator fidelity (Hoskins and El-Gindy 2006; Nilsson 1993), the specific driving task, and the realism of its implementation (Kaptein et al. 1996). Ultimately, literature shows that relative validity for driving simulators can be assumed, but absolute validity is limited (Mullen et al. 2011). This means that the validation results of the FRAM model are valid within the simulator environment but cannot be completely transferred to real on-road behaviour.

The calculation of the normal range of everyday variability per analysed function could be improved in the future by performing the reference scenarios 1–3 at least twice to discover which deviations in variability are normal, even if the participants are driving the same scenario again. However, this would increase the number of scenarios as well as the time needed, which was already high for the test subjects. This makes it a cost–benefit question, where we think that our simplified approach should be acceptable and sufficient.

In addition, the purely descriptive evaluation of the predictive validity can be criticised. It should be remembered that the focus of the predictive validity assessment was to analyse those functions, and how many functions, for which the predictions about performance variability through the FRAM model are valid or invalid rather than to know the number of test subjects for which the predictions are valid or not. The reason for this function focus is that potential invalid predictions could subsequently be refined to calibrate the model, which would otherwise be impossible. Therefore, it was not possible to calculate a distribution of the evaluation metrics per scenario, but only a single value in each case. This is why no inferential statistical analysis could be applied to evaluate the potential effects of changing environmental or human factors conditions.

Furthermore, scientific researchers can employ several statistical rates to evaluate binary classifications and their confusion matrices. In this work, the accuracy, HR, and CRR are used to evaluate the predictive validity in contrast to the MCC. This contradicts the general recommendation of Stanton and Young (1999a, 2003) to use the MCC as an appropriate statistic for the validation of human factors methods using the SDT, as well as the conclusion of Chicco and Jurman (2020) that the MCC is the most informative score for evaluating binary classification tasks and should be given preference over accuracy and F1 score by all scientific communities. However, the findings of Zhu (2020) challenge this general statement. Finally, there is no clear recommendation that just one specific metric should be used; this depends to a large extent on the context of the use and objective of the validation. Rather, a mix of different metrics, as applied in this paper, should be used to avoid misleading interpretations.

Last but not least, some methodological issues concerning the theoretical concept of the predictive validation approach can be identified. First, it is impossible to validate the whole FRAM model due to the overwhelming number of functions that have to be tested in a large and complex FRAM model. Only a few functions and their expected, as well as unexpected effects can be examined. Second, when manipulating one function, it is difficult to actually keep all of the remaining functions constant that were supposed to be constant, since the type of manipulation measure can potentially affect the performance of other functions. This problem is exacerbated by the fact that it is not even possible to check which functions this applies to, as it is impossible to analyse the performance variability for all functions. This results in interaction effects, whereby observed effects can no longer be fully attributed to the manipulated function. Furthermore, it might be difficult to find a targeted manipulation measure for each function in the model, e.g., for cognitive functions, since either no targeted manipulation is possible or several functions would be manipulated at the same time. Moreover, the extent and manner in which a manipulation has to be carried out to achieve the desired effect are generally unclear. Thus, following the method of constant stimuli from psychophysics (Fechner 1860), different stimulus intensities or types would have to be varied per manipulated function to see to which extent or manner a manipulation of an upstream function has to be carried out that results in a significant change in the performance variability of the individual downstream functions. Naturally, the extent and manner of the stimulus required vary between the individual downstream functions. Third, the performance variability of a downstream function may only change when several upstream inputs are varied instead of just the one manipulated function. Thus, an expected coupling could make sense and be valid even if no effect was observed in isolation. Consequently, all what-if combinations would have to be taken into account to be able to represent the complexity, which is simply impractical. Fourth, it is impossible to test whether there is also a direct influence for the functions that are indirectly influenced by the manipulated function. In addition, some functions are tested, where a direct influence by the manipulated function can be expected, and at the same time other functions that are also directly influenced by the manipulated function provide upstream inputs for the tested function. Hence, in these cases, there is always a degree of uncertainty as to whether the effect is direct or indirect.

5.3 Utility and applicability of the formal approach to assess predictive validity in FRAM

A research-practice gap of systemic models and methods (Underwood and Waterson 2012), especially FRAM, currently exists in literature, which means that researchers are presently applying systemic methods due to the current state-of-the-art and, in contrast, many practitioners press ahead with more traditional methods because of their ease of use or popularity despite known limitations (Grabbe et al 2022). Frequently mentioned reasons for this are a difficult and time-consuming application (Salmon et al. 2020), reduced model validation and usability, and a potential analyst bias (Underwood and Waterson 2012). Against this background, the results of the validation must be correlated to usability as a cost-effectiveness trade-off to be able to evaluate the utility benefit of the predictive validation approach in general (cf. Stanton and Young 2003). The effectiveness hereby represents the validity of the FRAM model to explain performance variability in an overtaking scenario, and the costs are related to the resources and time used by the method. As shown in Sect. 5.1, the validity is limited and can only be partly assumed. In contrast, the costs of using the method are high, since the model development by function identification and variability data collection was very time- and resource-consuming (Grabbe et al. 2022), something that also applies to the validation process. It should be noted that only two and not all of the functions of the model could be validated by this great effort. Therefore, the utility of the analysed FRAM model is questionable in terms of predictive validity if it is used as an evaluative method. On the other hand, the utility of the FRAM model as an analytical method is still an open question and difficult to demonstrate objectively.

In addition, and as shown in Sect. 5.2, there are several methodological issues related to the theoretical concept of the predictive validation approach for an FRAM model due to high complexity, leading to the conclusion that a complete validation of an FRAM model is impossible. Rather, the predictive validation approach developed in this paper should be applied to calibrate and not validate an FRAM model. This means that it can be used to select a few interesting functions in the model and to refine their modelling for a better understanding of their potential effects on the system behaviour with regard to specific system conditions, but not to prove that an FRAM model is valid or not. Consequently, the approach is appropriate to enhance any basic knowledge about system mechanisms gained by the FRAM model, but inappropriate to reach any final decisions concerning the approval of designs in safety–critical systems.

6 Conclusions and outlook

This paper developed a framework for evaluating the reliability and validity of an FRAM model, assessed the predictive validity of one specific FRAM model, and evaluated the applicability of this validation approach. The study shows that the validity and usefulness of the FRAM model by Grabbe et al. (2022) is limited and that the model results cannot be generalised to changing system conditions without any model adaptations. However, it is not clear whether this arises from the FRAM method itself or from the manner in which it was applied (cf. Stanton et al. 2013). Also, the applicability of the approach to demonstrate predictive validity is greatly reduced on account of several methodological limitations.

In future, the formal reliability and validity framework, and especially the predictive validation approach, should also be applied to other FRAM models in different application contexts so as to determine the reliability and validity generalisation of the FRAM method. Especially, human–machine dissonances and their predicted effects through an FRAM model should be validated. Moreover, the test–retest paradigm should be applied to rather small FRAM models to evaluate the reliability of the FRAM method and potential training effects in this context.

In conclusion, this paper contributes to making up for the lack of a formal validity approach for the FRAM method as well as to the research-practice gap of systemic HFE models and methods and their associated ongoing concerns of reliability and validity. In particular, this work helps analysts compare the cost-effectiveness of FRAM with other HFE methods. Overall, the developed framework provides a good foundation to evaluate the reliability and validity of an FRAM model. However, there is still potential for improvement and extension, especially against the background of the methodological advancement of FRAM and integration with other methods offering new opportunities for validation. Indeed, the reliability and validity framework can be used to calibrate rather than validate an FRAM model.

Acknowledgements The authors are grateful for all subjects who participated in this study. Furthermore, we would like to thank the editing service KERN for proofreading the paper. The authors are grateful to colleagues who provided valuable comments during the writing process of this paper.

Author contributions Conceptualisation, NG, AA; methodology, NG, AA; software, NG, AA; validation, NG, AA; formal analysis, NG, AA; investigation, NG, AA; resources, NG; data curation, NG, AA; writing—original draft preparation, NG; writing—review and editing, NG; visualisation, NG; supervision, KB; All authors have read and agreed to the published version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflicts of interest The authors have no relevant financial or non-financial interests to dis-close. In addition, no funds, grants or other support were received.

Institutional review board statement The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Institutional Review Board of the Ethics Committee of the Technical University of Munich (protocol code 509/21 S and date of approval 23 August 2021).

Informed consent Informed consent was obtained from all subjects involved in the study.

References

Accou B, Reniers G (2019) Developing a method to improve safety management systems based on accident investigations: the SAfety FRactal ANalysis. Saf Sci 115:285–293

Adriaensen A, Patriarca R, Smoker A, Bergström J (2019) A sociotechnical analysis of functional properties in a joint cognitive system: a case study in an aircraft cockpit. Ergonomics 62(12):1598–1616

Anfara VA Jr, Brown KM, Mangione TL (2002) Qualitative analysis on stage: making the research process more public. Educ Res 31(7):28–38

Annett J (2002) A note on the validity and reliability of ergonomics methods. Theor Issues Ergon Sci 3(2):228–232

Anvarifar F, Voorendt MZ, Zevenbergen C, Thissen W (2017) An application of the functional resonance analysis method (FRAM) to risk analysis of multifunctional flood defences in the Netherlands. Reliab Eng Syst Saf 158:130–141

Baber C, Stanton NA (1994) Task analysis for error identification: a methodology for designing error-tolerant consumer products. Ergonomics 37(11):1923–1941

Baber C, Stanton NA (1996) Human error identification techniques applied to public technology: predictions compared with observed use. Appl Ergon 27(2):119–131

Baber C, Young MS (2022) Making ergonomics accountable: reliability, validity and utility in ergonomics methods. Appl Ergon 98:103583

Balci O (1998) Verification, validation, and testing. Handb Simul 10(8):335–393

Banks J, Gerstein D, Searles SP (1987) Modeling processes, validation, and verification of complex simulations: a survey. In: 1987 SCS simulators conference, p 13–18

Baysari MT, Caponecchia C, McIntosh AS (2011) A reliability and usability study of TRACEr-RAV: the technique for the retrospective analysis of cognitive errors–for rail, Australian version. Appl Ergon 42(6):852–859

Blana E (1996) Driving simulator validation studies: a literature review. Working paper, Institute of Transport Studies, University of Leeds, Leeds, UK

Bridges KE, Corballis PM, Hollnagel E (2018) "Failure-to-Identify" hunting incidents: a resilience engineering approach. Hum Factors 60(2):141–159

Bulgren WG (1982) Discrete system simulation. Prentice Hall, Upper Saddle River

Chicco D, Jurman G (2020) The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genom 21(1):1–13

Cornelissen M, McClure R, Salmon PM, Stanton NA (2014) Validating the strategies analysis diagram: assessing the reliability and validity of a formative method. Appl Ergon 45(6):1484–1494

Crawford A (1963) The overtaking driver. Ergonomics 6(2):153–170

Creswell JW, Miller DL (2000) Determining validity in qualitative inquiry. Theory Pract 39(3):124–130

Dallat C, Salmon PM, Goode N (2017) Risky systems versus risky people: to what extent do risk assessment methods consider the systems approach to accident causation? A review of the literature. Saf Sci 119:266–279

Espié S, Gauriat P, Duraz M (2005) Driving simulators validation: the issue of transferability of results acquired on simulator. In: Driving simulation conference North-America (DSC-NA 2005), Orlondo, FL

Fechner GT (1860) Elemente der Psychophysik [elements of psychophysics]. Breitkopf und Härtel, Leipzig, pp 280–286

Ferreira PN, Cañas JJ (2019) Assessing operational impacts of automation using functional resonance analysis method. Cognit Technol Work 21:1–18

Goode N, Salmon PM, Taylor NZ, Lenné MG, Finch CF (2017) Developing a contributing factor classification scheme for Rasmussen's AcciMap: reliability and validity evaluation. Appl Ergon 64:14–26

Grabbe N, Kellnberger A, Aydin B, Bengler K (2020) Safety of automated driving: the need for a systems approach and application of the functional resonance analysis method. Saf Sci 126:104665

Grabbe N, Gales A, Höcher M, Bengler K (2022) Functional resonance analysis in an overtaking situation in road traffic: comparing the performance variability mechanisms between human and automation. Safety 8(1):3

Hanssmann F (2018) Einführung in die Systemforschung. Oldenbourg Wissenschaftsverlag, Munich

Harris D, Stanton NA, Marshall A, Young MS, Demagalski J, Salmon P (2005) Using SHERPA to predict design-induced error on the flight deck. Aerosp Sci Technol 9(6):525–532

Hill R, Boult M, Sujan M, Hollnagel E, Slater D (2020) Predictive analysis of complex systems' behaviour. https://www.researchgate.net/profile/David-Slater/publication/343944100_PREDICTIVE_ANALYSIS_OF_COMPLEX_SYSTEMS'_BEHAVIOUR_SWIFTFRAM/links/5f4907e0299bf13c5047f8d3/PREDICTIVE-ANALYSIS-OF-COMPLEX-SYSTEMS-BEHAVIOUR-SWIFTFRAM.pdf. Accessed 18 Nov 2021

Hollnagel E (2004) Barriers and accident prevention. Ashgate, Hampshire

Hollnagel E (2012) FRAM: the functional resonance analysis method: modelling complex socio-technical systems. CRC Press, Boca Raton

Hollnagel E (2014) Safety–I and safety–II: the past and future of safety management. CRC Press, Boca Raton

Hollnagel E (2020) FRAM model interpreter. https://functionalresonance.com/onewebmedia/FMI%20basicPlus%20V3.pdf. 09 Nov 2021

Hollnagel E, Hounsgaard J, Colligan L (2014) FRAM—the functional resonance analysis method—a handbook for the practical use of the method. https://functionalresonance.com/onewebmedia/FRAM_handbook_web-2.pdf. 17 Nov 2021

Hoskins AH, El-Gindy M (2006) Technical report: Literature survey on driving simulator validation studies. Int J Heavy Veh Syst 13(3):241–252

Hughes BP, Newstead S, Anund A, Shu CC, Falkmer T (2015) A review of models relevant to road safety. Accid Anal Prev 74:250–270

Hulme A, Stanton NA, Walker GH, Waterson P, Salmon PM (2019) What do applications of systems thinking accident analysis methods tell us about accident causation? A systematic review of applications between 1990 and 2018. Saf Sci 117:164–183

Hulme A, Stanton NA, Walker GH, Waterson P, Salmon PM (2021a) Testing the reliability and validity of Net-HARMS: a new systems-based risk assessment method in HFE. In: Congress of the International Ergonomics Association, Springer, Cham, p 354–362

Hulme A, Stanton NA, Walker GH, Waterson P, Salmon PM (2021b) Testing the reliability and validity of risk assessment methods in Human Factors and Ergonomics. Ergonomics 65:1–22

Hulme A, Stanton NA, Walker GH, Waterson P, Salmon PM (2021c) Are accident analysis methods fit for purpose? Testing the criterion-referenced concurrent validity of AcciMap, STAMP-CAST and AcciNet. Saf Sci 144:105454

ISO Standard 26262 (2018) Road vehicles—functional safety—part 3: concept phase. https://www.iso.org/standard/68385.html. Accessed 16 Dec 2021

Jensen A, Aven T (2018) A new definition of complexity in a risk analysis setting. Reliab Eng Syst Saf 171:169–173

Kaptein NA, Theeuwes J, Van Der Horst R (1996) Driving simulator validity: some considerations. Transp Res Rec 1550(1):30–36

Kaya GK, Ovali HF, Ozturk F (2019) Using the functional resonance analysis method on the drug administration process to assess performance variability. Saf Sci 118:835–840

Kirwan B, Kennedy R, Taylor-Adams S, Lambert B (1997) The validation of three human reliability quantification techniques—THERP, HEART and JHEDI: Part II—results of validation exercise. Appl Ergon 28(1):17–25

Laaraj N, Jawab F (2018) Road accident modeling approaches: literature review. In: 2018 International colloquium on Lo-1769 gistics and supply chain management (LOGISTIQUA). IEEE, p 188–193

Larsson P, Dekker SW, Tingvall C (2010) The need for a systems theory approach to road safety. Saf Sci 48(9):1167–1174

Larue GS, Wullems C, Sheldrake M, Rakotonirainy A (2018) Validation of a driving simulator study on driver behavior at passive rail level crossings. Hum Factors 60(6):743–754

Leveson N (2004) A new accident model for engineering safer systems. Saf Sci 42(4):237–270

Li W, He M, Sun Y, Cao Q (2019) A proactive operational risk identification and analysis framework based on the integration of ACAT and FRAM. Reliab Eng Syst Saf 186:101–109

Liebl F (2018) Simulation. Oldenbourg Wissenschaftsverlag, Munich

MacKinnon RJ, Pukk-Härenstam K, Kennedy C, Hollnagel E, Slater D (2021) A novel approach to explore safety-I and safety-II perspectives in in situ simulations—the structured what if functional resonance analysis methodology. Adv Simul 6(1):1–13

Makeham MA, Stromer S, Bridges-Webb C, Mira M, Saltman DC, Cooper C, Kidd MR (2008) Patient safety events reported in general practice: a taxonomy. BMJ Qual Saf 17(1):53–57

Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim Biophys Acta (BBA) Protein Struct 405(2):442–451

Mullen N, Charlton J, Devlin A, Bedard M (2011) Simulator validity: behaviours observed on the simulator and on the road. In: Fisher DL, Rizzo M, Caird JK, Lee JD (eds) Handbook of driving simulation for engineering, medicine and psychology, 1st edn. CRC Press, Boca Raton, pp 1–18

Nemeth C (2013) Erik Hollnagel: FRAM: the functional resonance analysis method, modeling complex socio-technical systems. Cogn Technol Work 1(15):117–118

Nilsson L (1993) Behavioural research in an advanced driving simulator-experiences of the VTI system. In: Proceedings of the human factors and ergonomics society annual meeting, vol 37, no 9. Sage: Los Angeles, p 612–616

O'Connor P (2008) HFACS with an additional layer of granularity: validity and utility in accident analysis. Aviat Space Environ Med 79(6):599–606

Olsen NS (2013) Reliability studies of incident coding systems in high hazard industries: a narrative review of study methodology. Appl Ergon 44(2):175–184

Olsen NS, Shorrock ST (2010) Evaluation of the HFACS-ADF safety classification system: inter-coder consensus and intra-coder consistency. Accid Anal Prev 42(2):437–444

Patriarca R, Bergström J (2017) Modelling complexity in everyday operations: functional resonance in maritime mooring at quay. Cogn Technol Work 19(4):711–729

Patriarca R, Bergström J, Di Gravio G (2017) Defining the functional resonance analysis space: combining abstraction hierarchy and FRAM. Reliab Eng Syst Saf 165:34–46

Patriarca R, Di Gravio G, Woltjer R, Costantino F, Praetorius G, Ferreira P, Hollnagel E (2020) Framing the FRAM: a literature review on the functional resonance analysis method. Saf Sci 129:104827

Pereira AG (2013) Introduction to the Use of FRAM on the effectiveness assessment of a radiopharmaceutical dispatches process. In: International nuclear Atlantic conference

Pollatschek M, Polus A (2005) Modelling impatience of driver in passing manuevers. Transp Traffic Theory 16:267–279

Qureshi ZH (2007) A review of accident modelling approaches for complex socio-technical systems. In: Proceedings of the 1757 twelfth Australian workshop on Safety critical systems and software and safety-related programmable systems, vol 86. Australian Computer Society, Inc., p 1758 47–59

Rasmussen J (1997) Risk management in a dynamic society: a modelling problem. Saf Sci 27(2–3):183–213

Rastegary H, Landy FJ (1993) The interactions among time urgency, uncertainty, and time pressure. In: Time pressure and stress in human judgment and decision making. Springer, Boston, p 217–239

Ross A, Sherriff A, Kidd J, Gnich W, Anderson J, Deas L, Macpherson L (2018) A systems approach using the functional resonance analysis method to support fluoride varnish application for children attending general dental practice. Appl Ergon 68:294–303

Salehi V, Veitch B, Smith D (2021) Modeling complex socio-technical systems using the FRAM: a literature review. Hum Factors Ergon Manuf Serv Ind 31(1):118–142

Salmon PM, McClure R, Stanton NA (2012) Road transport in drift? Applying contemporary systems thinking to road safety. Saf Sci 50(9):1829–1838

Salmon PM, Read GJ, Walker GH, Stevens NJ, Hulme A, McLean S, Stanton NA (2020) Methodological issues in systems Human Factors and Ergonomics: perspectives on the research–practice gap, reliability and validity, and prediction. Hum Factors Ergon Manuf Serv Ind. https://doi.org/10.1002/hfm.20873

Sargent RG (1984) A tutorial on verification and validation of simulation models. In: Proceedings of the 16th conference on Winter simulation, pp 115–121. https://repository.lib.ncsu.edu/bitstream/handle/1840.4/4929/1984_0017.pdf?sequence=1

Schrank WE, Holt CC (1967) Critique of: "Verification of computer simulation models." Manag Sci 14(2):B-104

Sheridan TB (1992) Telerobotics, automation, and human supervisory control. MIT Press, Cambridge

Stanton NA (2014) Commentary on the paper by Heimrich Kanis entitled 'Reliability and validity of findings in ergonomics research': where is the methodology in ergonomics methods? Theor Issues Ergon Sci 15(1):55–61

Stanton NA (2016) On the reliability and validity of, and training in, ergonomics methods: a challenge revisited. Theor Issues Ergon Sci 17(4):345–353

Stanton NA, Baber C (2005) Validating task analysis for error identification: reliability and validity of a human error prediction technique. Ergonomics 48(9):1097–1113

Stanton NA, Stevenage SV (1998) Learning to predict human error: issues of acceptability, reliability and validity. Ergonomics 41(11):1737–1756

Stanton NA, Young MS (1999a) What price ergonomics? Nature 399(6733):197–198

Stanton NA, Young MS (1999b) A guide to methodology in ergonomics: designing for human use. Taylor & Francis, London

Stanton NA, Young MS (2003) Giving ergonomics away? The application of ergonomics methods by novices. Appl Ergon 34(5):479–490

Stanton NA, Salmon P, Harris D, Marshall A, Demagalski J, Young MS, Dekker S et al (2009) Predicting pilot error: testing a new methodology and a multi-methods and analysts approach. Appl Ergon 40(3):464–471

Stanton NA, Salmon PM, Rafferty LA, Walker GH, Baber C, Jenkins DP (2013) Human factors methods: a practical guide for engineering and design. CRC Press, Boca Raton

Stanton NA, Brown JW, Revell KM, Clark JR, Richardson J, Langdon P et al (2021a) Modelling automation-human driver interactions in vehicle takeovers using OESDs. Designing interaction and interfaces for automated vehicles. CRS Press, Boca Raton, pp 299–320

Stanton NA, Brown JW, Revell KM, Kim J, Richardson J, Langdon P et al (2021b) OESDs in an on-road study of semi-automated vehicle to human driver handovers. Cognit Technol Work 24:1–16

Stanton NA, Brown JW, Revell KM, Kim J, Richardson J, Langdon P et al (2021c) Validating OESDs in an on-road study of semi-automated vehicle-to-human driver takeovers. In: Designing interaction and interfaces for automated vehicles. CRC Press, Boca Raton, p 443–464b

Stanton NA, Brown JW, Revell KM, Langdon P, Bradley M, Politis I et al (2021d) Validating operator event sequence diagrams: the case of automated vehicle-to-human driver takeovers. In: Designing interaction and interfaces for automated vehicles. CRC Press, Boca Raton, p 137–157

Tapio J (2003) Ohitukset kaksikaistaisilla teilla (Summary in English). Finnish Road Administration, Helsinki

Underwood P, Waterson P (2012) A critical review of the STAMP, FRAM and Accimap systemic accident analysis models. In: Advances in human aspects of road and rail transportation. CRC Press, Boca Raton, pp 385–394

Van Horn RL (1971) Validation of simulation results. Manag Sci 17(5):247–258

Vanderhaegen F (2014) Dissonance engineering: a new challenge to analyse risky knowledge when using a system. Int J Comput Commun Control 9(6):776–785

Vanderhaegen F (2016) A rule-based support system for dissonance discovery and control applied to car driving. Expert Syst Appl 65:361–371

Vanderhaegen F (2021) Heuristic-based method for conflict discovery of shared control between humans and autonomous systems—a driving automation case study. Robot Auton Syst 146:103867

Wege CA, Pereira M, Victor TW, Krems JF, Stevens A, Brusque C (2014) Behavioural adaptation in response to driving assistance technologies: a literature review. Driver adaptation to information and assistance systems. The Institution of Engineering and Technology, London, p S.3-34

Wienen HCA, Bukhsh FA, Vriezekolk E, Wieringa RJ (2017) Accident analysis methods and models—a systematic literature 1767 review. In: Centre for Telematics and Information Technology (CTIT), p 1768

Woltjer R, Hollnagel E (2008) Functional modeling for risk assessment of automation in a changing air traffic management environment. In: Proceedings of the 4th international conference working on safety, vol 30

Yang Q, Tian J, Zhao T (2017) Safety is an emergent property: illustrating functional resonance in air traffic management with formal verification. Saf Sci 93:162–177

Yang Q, Tian J (2015) Model-based safety assessment using FRAM for complex systems. In: Proceedings of the 25th European safety and reliability conference

Zhu Q (2020) On the performance of Matthews correlation coefficient (MCC) for imbalanced dataset. Pattern Recognit Lett 136:71–80

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

F Complementary Publications

“Once you stop learning, you start dying.”
– Albert Einstein –

Albers, D., Grabbe, N., Forster, Y., Naujoks, F., Keinath, A., & Bengler, K. (2022). (Don’t) Talk to Me! Application of the Kano Method for Speech Outputs in Conditionally Automated Driving. In Human Factors in Transportation (pp. 508-515).

Albers, D., Grabbe, N., Janetzko, D., & Bengler, K. (2020, September). Saluton! How do you evaluate usability?–Virtual Workshop on Usability Assessments of Automated Driving Systems. In 12th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 109-112).

Albers, D., Radlmayr, J., Grabbe, N., Hergeth, S., Naujoks, F., Forster, Y., ... & Bengler, K. (2021, May). Human-machine interfaces for automated driving: Development of an experimental design for evaluating usability. In Proceedings of the 21st Congress of the International Ergonomics Association (IEA 2021) Volume III: Sector Based Ergonomics (pp. 541-551). Cham: Springer International Publishing.

Josten, J., Seewald, P., Eckstein, L., Hiller, J., Dautzenberg, P., Becker, D., (..), Hübner, M. (2023). Level 2 hands-off –Recommendations and guidance. FAT-Schriftenreihe 369. Forschungsvereinigung Automobiltechnik e.V. (FAT). Retrieved from: https://www.vda.de/de/aktuelles/publikationen/publication/level-2-hands-o---recommendations-and-guidance#publication-title.

Keler, A., Malcolm, P., Grigoropoulos, G., & Grabbe, N. (2020, October). Extraction and analysis of massive skeletal information from video data of crowded urban locations for understanding implicit gestures of road users. In 2020 IEEE Intelligent Vehicles Symposium (IV) (pp. 101-106). IEEE.

G Supervised theses

“The objective of education is learning, not teaching.”
– Russel Ackoff –

Bachelor’s theses:

Arifagic, A. (2021). Validation of a FRAM model in a Driving simulator study: The case of overtaking on a rural road (Bachelor’s thesis, Technical University of Munich).

Aydin, B. (2019). Risk Assessment of Automated Driving - Applicability of the Functional Resonance Analysis Method (Bachelor’s thesis, Technical University of Munich).

Fischer, L. (2019). Use of explicit and implicit communication of pedestrians in road traffic as groundwork for algorithms of vehicle automation (Bachelor’s thesis, Technical University of Munich).

Gales, A. (2021). Evaluation of road safety between a human driver and a fully automated vehicle: A socio-technical modelling approach using FRAM (Bachelor’s thesis, Technical University of Munich).

Kellnberger, A. (2019). Methods and Models to Define the Contribution of the Human Driver to Road Traffic Safety (Bachelor’s thesis, Technical University of Munich).

Krampert, L. (2018). Touch interaction in vehicle interior – benchmark and design recommendations for in-vehicle infotainment systems based on touch concepts (Bachelor’s thesis, Technical University of Munich).

Lieblein, C. (2022). The effect of the distance between two pedestrians on the crossing decision during the interaction with an automated vehicle - a VR user study (Bachelor’s thesis, Technical University of Munich).

Miholich, M. (2019). Performance of current driver assistance systems and vehicle automation (Bachelor’s thesis, Technical University of Munich).

Semester’s theses:

Artner, M. (2021). Literature Research regarding objective Tools and Methods for Pilots‘ Evaluation Procedures (Semester’s thesis, Technical University of Munich).

Eikam, N. (2020). Analysis of the behavior of children as pedestrians in road traffic and derivation of requirements for the interaction with an automated vehicle (Semester’s thesis, Technical University of Munich).

Foth, F. (2020). Design of a Ground Control Station for Planning and Monitoring of Range and Mission of Unmanned Aircrafts (Semester’s thesis, Technical University of Munich).

Grundmüller, V. (2020). Overview of current approaches to the safety assessment of automated vehicles and their potential estimations (Semester’s thesis, Technical University of Munich).

Höcher, M. (2019). Usefulness and Use Cases of Vehicle Automation in the Road Traffic System (Semester’s thesis, Technical University of Munich).

Lintner, T. (2020). Overview of Theories, Methods, and Models for Evaluating Acceptance and Perceived Risk in the Context of Automated Driving (Semester’s thesis, Technical University of Munich).

Ohme, C. (2019). Methods and metrics for behavioral analysis of vulnerable road users (Semester’s thesis, Technical University of Munich).

Plötz, E. (2019). Systemic approach to the safety analysis of the decision and communication behaviour of an Airbus A320 cockpit crew during an emergency situation (Semester’s thesis, Technical University of Munich).

Thanos, A. (2019). Opportunities and Risks of Vehicle Automation – Derivation of Relevant Application and Test Scenarios (Semester’s thesis, Technical University of Munich).

Master’s theses:

Batschkus, D. (2019). Concept, Development, and Initial Validation of an Algorithm for the Determination of Visual Attention Allocation to Dynamic Areas of Interest (Master’s thesis, Technical University of Munich).

Chen, L. (2023). Modeling of the interaction between pedestrians and driver or automated vehicles in complex multi-agent scenarios (Master’s thesis, Technical University of Munich).

Harre, R. (2019). Systemic approach to the safety analysis of decision and communication behavior of a cockpit crew of an Airbus 340-600 during an emergency situation (Master’s thesis, Technical University of Munich).

Höcher, M. (2019). Method for identifying and modelling the functions of a FRAM model in road traffic (Master’s thesis, Technical University of Munich).

Kirmayer, S. (2021). Human-Centered Design – How it is imagined vs. how it is done? (Master’s thesis, Technical University of Munich).

Krug, T. (2019). User-centric Evaluation of Connected Car Services regarding User Experience (Master’s thesis, Technical University of Munich).

Lintner, T. (2021). Potential of Smartphones in the Interaction Between Distracted Pedestrians and Automated Vehicles (Master’s thesis, Technical University of Munich).

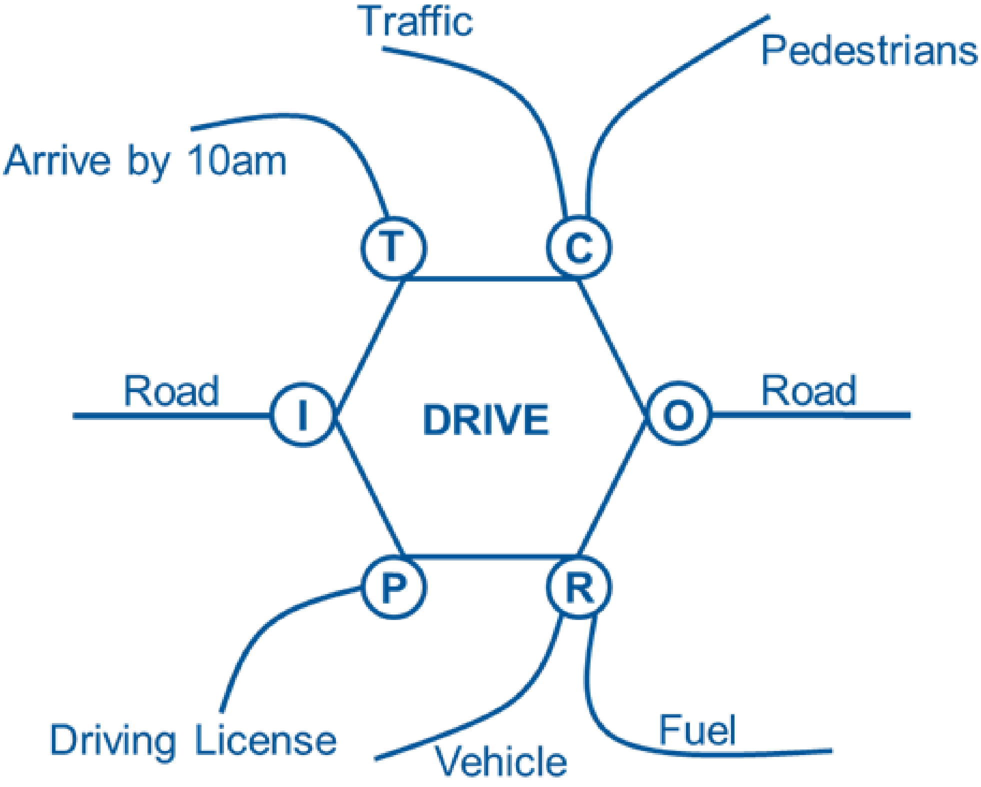
Mangold, U. (2020). Analysis and Modeling of the Interaction between Pedestrians and Drivers – Derivation of Requirements for the Automated Vehicle (Master’s thesis, Technical University of Munich).

Nguyen, L. (2019). Design of a pedestrian-driver-interaction model to support the development of automated vehicles in the city (Master’s thesis, Technical University of Munich).

Rosner, M. (2021). Analysis of the Pilot Ratings Generated in Training Operations to Derive Individual Training Requirements (Master’s thesis, Technical University of Munich).

Schmidl, P. (2019). Model-Based Manipulation of Driver Behavior Models for the Purpose of Generating Simulated Skid-Scenarios (Master’s thesis, Technical University of Munich).

Voß, J. (2019). Spatial Disorientation – Definition and Design of a Cognitive Assistance System for Pilot Support (Master’s thesis, Technical University of Munich).


Traffic
Pedestrians
Arrive by 10am
T
C
Road
I
DRIVE
O
Road
P
R
Driving License
Vehicle
Fuel