CONCEPTUALISING AUTOMATED DRIVING SHARED CONTROL HAZARD CAUSES

HELEN E. MONKHOUSE

Doctor of Philosophy (PhD)

University of York

Computer Science

August 2022

The most difficult thing is the decision to act.
The rest is merely tenacity.

— *Amelia Earhart*

Live as if you were to die tomorrow.
Learn as if you were to live forever.

— *Mahatma Gandhi*

Helen Elizabeth Monkhouse BEng(Hons) CEng MIET MWES: *Conceptualising automated driving shared control hazard causes,*
PhD Research Thesis © August 2022

SUPERVISORS:
John A McDermid
Ibrahim Habli

ABSTRACT

The motivation for this research was the realisation that the introduction of greater vehicle automation would not only change the task of driving but would also potentially change how vehicles are developed and safety is assured. Undertaking a practice-based workshop identified many Automated Driving (AD) safety assurance challenges having different levels of human-machine control. These challenges include an increase in the size and complexity of AD safety analyses, a need to re-examine the notion of controllability in the context of shared control, and the need to conceptualise the vehicle as a system of systems.

To begin addressing these challenges and to answer the research question "how can the safety of AD be assured under different levels of shared control?" this research has created three products: a vehicle model and behavioural competency taxonomy that allows AD shared control to be conceptualised, a concrete hazard analysis method for analysing AD shared control hazard causes, and a safety case argument pattern for that.

A series of case studies evaluate the research products described above. These cases have used contemporary AD vehicle features, having varying levels of automation. The evaluation of driver assistance, partial and conditional automation cases have been completed by the author. Complementing these is the analysis of a highly automated vehicle system, which has been undertaken with the engineering team from Oxbotica. Considered together these case studies establish the research products as a proof-of-concept hazard analysis method for AD shared control. Further evaluation work is needed to test the viability of the method as an engineering tool for use by automotive practitioners working in a product development environment.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Wow! This has been a 'journey'...

Thank you to my supervisors, Professor John McDermid and Professor Ibrahim Habli, for providing guidance and feedback throughout this research degree. Thank you to my family and friends, for putting up with my aloofness, for providing guidance and for being a sounding board when required. Thanks also to my employer, HORIBA MIRA Ltd., for affording me some flexibility to juggle my time between the 'day job' and a research degree.

A special thank you must go to Dr. Frances Finn, Sophie Berreen and Jackie Walsh who set me straight each and every time I threatened to quit.

Thank goodness I listened to you, lovely ladies!

PUBLICATIONS

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References. Parts of this work have been published in conference proceedings and journals previously. The details of which are shown below:

- John Birch, Roger Rivett, Ibrahim Habli, Ben Bradshaw, John Botham, Dave Higham, Peter Jesty, **Helen Monkhouse** and Robert Palin, "Safety cases and their role in ISO 26262 functional safety assessment", Computer Safety, Reliability, and Security, Springer, 2013.
- **Helen Monkhouse**, Ibrahim Habli and John McDermid, "The Notion of Controllability in an Autonomous Vehicle Context", CARS, 2015.
- **Helen Monkhouse**, Ibrahim Habli, John McDermid, Siddartha Khastgir and Gunwant Dhadyalla, "Why Functional Safety Experts Worry About Vehicle Systems Having Greater Autonomy," ATC 2017.
- John Birch, David Blackburn, John Botham, Ibrahim Habli, David Higham, **Helen Monkhouse**, Gareth Price, Norina Ratiu and Roger Rivett, "A Structured Argument for Assuring Safety of the Intended Functionality (SOTIF)," SAFECOMP, 2020.
- **Helen Monkhouse**, Ibrahim Habli and John McDermid, "An Enhanced Vehicle Control Model for Assessing Highly Automated Driving Safety," Reliability Engineering and System Safety, 2020.

Part I

INTRODUCTION AND PROBLEM DEFINITION

Driving a motor vehicle is an inherently dangerous pursuit. Although road safety has improved significantly over the years, estimates suggest that "traffic fatalities" will rank as the #5 cause of death by 2030 [201]. This grim statistic motivates National and International initiatives seeking to reduce traffic fatalities and serious injuries to zero. Central to these initiatives is the introduction of greater automation.

AD has the potential to revolutionise the relationship that society has with vehicles and road transport. Achieved by reducing the number of accidents caused by human error, but also by potentially reducing pollution, easing congestion, preserving the mobility of an ageing population, and increasing overall mobility efficiency. However, automation is changing the types and causes of hazards, and influencing the way in which users interact with complex systems, particularly challenging the notion of human control as a primary basis for hazard mitigation. This necessitates the re-evaluation of the driver-in-the-loop model to extend the notion of controllability to the shared control of highly automated tasks.

This assault on the automotive functional safety paradigm is the motivation behind this research. A description of the research contributions concludes Part I.

1

ROAD SAFETY RESEARCH

Travelling in an early 20th Century motor vehicle was a dangerous pursuit. With vehicles being heavy and difficult to control, serious accidents were common. Early road safety research was focused on the driver, with 1920's road traffic accident prevention focusing on the identification and punishment of the "accident prone driver". That view perpetuated well into the 1950's. This was due in part to Greenwood and Yule's 1920's research into World War I munitions factory injuries, and Sigmund Freud's psychological testing to identify accident proneness [46].

Road safety research has evolved through a number of phases: From exploring *what* is happening during an accident in the hope of answering *why* accidents happen. To then applying systems thinking in an attempt to understand *how* accidents happen. Through to contemporary road safety research that attempts to understand the traffic system as a *whole*; resulting in complex and multi-disciplined accident models [139].

Seminal research by Haddon, Suchman and Klein developed a framework to describe the phases of a traffic accident and its contributing factors [67], which moved road safety from a blame to a prevention focused culture. While Haddon, Suchman and Klein's research drove improvements in vehicle safety, 1960 road safety research still saw significant efforts focused on identifying, grouping and "removing" (via educational programmes) the "accident prone driver" [139]. However, with research[1] suggesting that less

1 An Israeli study comparing road traffic accidents with driver demographic over a 20 year period suggests that the accident distribution follows an exponential decay: With 82.6% of drivers having had only one accident during the period from 1983 to 2004. 13.4% having had 2 accidents in the same time period. While drivers having 3 accidents or more accounted for less than 5% of half a million or so accidents [54].

than 5% of drivers have multiple accidents, the notion of the "accident prone driver" is unfounded.

1.1 ROAD DESIGN SAFETY

Since the early 20th Century road design safety has been influenced by standards and design guidelines. Early standards and guidelines focused on achieving correct sighting distances; determined from typical driver reaction times, vehicle speed and performance. However, Shalom Hakkert and Gitelman make the observation that because the effectiveness of these standards and guidelines at reducing accident rates was not measured, important road characteristics affecting safety were not addressed [139]. For example, the relative position of roadside furniture at junctions, road camber and lighting all have the potential to influence accident rates and severity but did not form part of the early guidance. Contemporary road safety standards and guidelines[2] have an empirical foundation and address the potential impacts of the various road design attributes during the accident phases. For example, the design of junctions and roundabouts to reduce the likelihood of accidents [46], and the use of energy absorbent roadside structures that reduce accident severity by making impacts more "forgiving" [32].

2 For example, the US Highway Safety Manual [160] and the World Road Association's Road Safety Manual [202].

1.2 VEHICLE DESIGN SAFETY

1.2.1 *Vehicle design features*

Although the accident statistics[3] do highlight the significant risk associated with travelling by car at the early 1900's, these early motor vehicles did include a number of mechanical safety measures. Early mechanical safety devices included wiper blades (1903), the rear view mirror (1911), the laminated windscreen (1927), speedometer (1937) and padded dashboards (1947). Safety features, such as airbags (1951), disk brakes (1953), and the three-point seatbelt (1959), were introduced through the 1950's and became prevalent once legislation mandated their use [12, 75].

The use of electronics in vehicle safety systems has increased steadily since the 1980's, with Mercedes-Benz offering the first Secondary Restraints System (SRS) in 1981 and together with BMW and Toyota, offering traction control in 1987 [12].

Restraints systems are typically referred to as *passive safety* systems because they reduce the severity of the resulting accident. The first *active safety* systems appeared during the 1990's. These systems actively support the driver to prevent a critical situation becoming an accident. Examples of *active safety* systems include: Anti-lock Braking System (ABS) and Electronic Stability Control (ESC) that make control inputs to the vehicle's hydraulic braking system that override those of the driver. By taking over braking authority, when the vehicle is deemed to have become unstable, these systems support the driver to maintain control during critical situations thus avoiding an accident.

Complex electronic control systems are now pervasive within the vehicle, with today's systems providing *tactical safety* support to the driver. That is, supporting the driver with Advanced Driver Assistance Systems (ADASs) that

3 In a historical discussion about road safety improvements the UK Driver & Vehicle Standards Agency states that in 1931, when the Highway Code was introduced in the UK, 7,343 people were killed when only 2.4 million vehicles were in use. In contrast, in 2008, 2,538 people were killed with 26.5 million vehicles on the road [75].

aim to prevent a normal driving situation becoming critical. Example ADAS features include Automatic Emergency Braking (AEB), Blind Spot Monitoring (BSM) and Lane Keeping Assistance (LKA). The impact of automation on the task of driving is discussed in more detail in Chapter 2.

1.2.2 *Functional safety*

The first embedded electronic control systems to appear preceded the *passive safety* systems described above. These systems first appeared in passenger vehicles in the 1980's and were introduced to replace specific mechanical engine controls (e.g. electronic ignition, fuel injection). A drive for greater efficiency and the introduction of more stringent emissions targets, meant that within a decade everything from driver demand to air and fuel mix were under the control of embedded programmable electronics.

From their introduction in the early 1980's, concern about the impact that a control system failure might have on vehicle safety grew, until in the early 1990's the UK government set up the Motor Industry Software Reliability Association (MISRA) consortium, with the objective to "provide assistance to the automotive industry in the creation and application within a vehicle system of safe, reliable software". The "MISRA Guidelines" [112], first published in 1994, provided the practising engineer with guidance describing the activities necessary to evaluate the safety implications of programmable automotive systems, and to achieve *functional safety*; that is, the absence of unreasonable risk caused by embedded electronics hardware and software faults. This included a method for assessing risk (the MISRA Risk Model) that considered the effects of failure on system behaviour (the hazards) and used, amongst other things, the concept of controllability (i.e. the ability of 'persons in harm's way' to make the correct and timely reaction to avoid the harm) to estimate the degree of risk associated with vehicle hazards.

Since its introduction in the mid-1990s the concept of controllability has been widely adopted by industry [109]. In 2011 the automotive standard ISO 26262

[162] was released and continues the philosophies established by MISRA by including a risk classification scheme, which although slightly different from the original MISRA risk scheme still has the concept of controllability at its heart.

Although not explicitly stated in ISO 26262 [162], the controllability concept used to classify hazards characterised by vehicle movement control, makes three assumptions about the driver: Firstly, it assumes that the driver is always part of the control loop. This assumption originates from the "Convention on Road Traffic", which is commonly referred to as the 'Vienna Convention'[4] [178]. Secondly, that the driver is integral to control, as illustrated by the Vehicle Control Model (VCM) in Figure 1, implying that the driver is fully aware of their surroundings, i.e. achieving full situational awareness. The third assumption is that the *'safe state'*, entered by the system following a catastrophic failure, can be to simply shutdown the system; i.e. to *'fail safe'* or to *'fail passive'*. The introduction of ever more complex AD features and the move towards highly automated and connected driving changes these driver assumptions. Not only is the driver's role and behaviour potentially changing [30, 151], but also vehicles are becoming part of a transport system of systems; with the vehicle potentially modifying its behaviour as a result of information received from other vehicles, from roadside infrastructure, and from the Cloud. The potential impact that this might have on the driver is explored further in Chapter 2.

1.3 A SYSTEMS SAFETY APPROACH

As discussed at the beginning of this chapter, Haddon, Suchman and Klein's research was pivotal in a multi-faceted system-led approach to road safety, which included considering the phases of an accident (i.e. pre-crash, crash and post-crash). Represented as a two-dimensional matrix, the Haddon Matrix [197] captures the different factors and design interventions (i.e. human factors,

4 The Vienna Convention is an international treaty which by establishing 'standard traffic rules' for such things as vehicles, road signage and driver obligations, seeks to facilitate cross-boarder traffic movement and improve road safety [178].

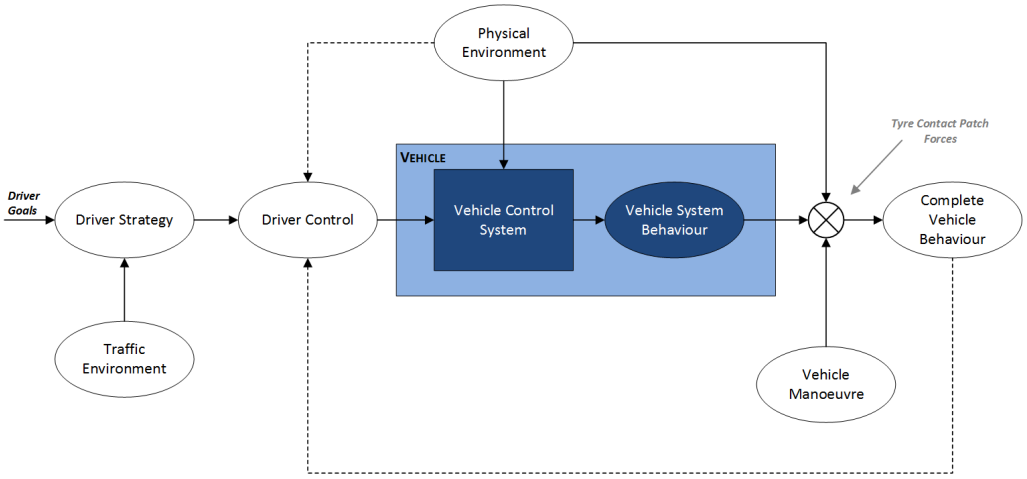


Figure 1: Driver in the loop – "control systems view of the vehicle" (reproduced from [113])

vehicle and environmental) that can be attributed to the accident phases. As Shalom Hakkert and Gitelman highlights, Haddon used analogies from other domains to communicate the need for a preventative approach to road safety. For example, installing non-slip flooring surfaces rather than telling employees to walk more carefully. Or designing guillotines that require two-handed operation, to prevent limb entrapment, rather than sending the operator on a training course [139]!

Contemporary road traffic research seeks to understand accidents in a multi-disciplinary way; looking at the plethora of circumstances leading up to an accident, thus identifying countermeasures to avoid future accidents. Although one might imagine cost being a critical factor in determining which countermeasures to deploy, Shalom Hakkert and Gitelman suggests that law enforcement does not always provide the most cost-effective countermeasure (e.g. speed limits and speed enforcement cameras). However, they suggest that these countermeasures are still used extensively because speed limits and speed enforcement may establish and maintain the *normative* driver behaviour [139].

1.4 REGULATION AND CONSUMER DRIVEN SAFETY

Although some vehicle safety related legislation may have been in place at national or regional levels much earlier, the 1960's saw the consolidation of safety regulations and the development of a legal framework, in Europe, the Rest of the World and in the United States. The "UNECE World Forum Harmonization of Vehicle Regulations (WP.29)" regulatory forum was established, with the 1958 Agreement establishing requirements for vehicles, vehicle systems, parts and equipment; including performance requirements, conformity of production, and in service inspections [177, 198].

In the United States, motivated by high vehicle related fatalities,[5] the "National Traffic and Motor Vehicle Safety Act" was passed in 1966. The act provided the legal framework enabling the Federal Government to develop and implement vehicle and highway safety standards [193].

Seen by many in the automotive industry as representing the 'minimum' performance requirement, automotive manufacturers own product design requirements typically exceed those specified by the regulations. A number of high profile incidents, relating to Mini power steering system failures, illustrates the need to potentially engineer above and beyond the regulatory requirements[6]. With vehicle safety having the potential to act as a 'brand differentiator', while helping to reduce the cost of ownership (e.g. annual insurance premiums), vehicle safety features undoubtably have commercial

5 While signing the 1966 "National Traffic and Motor Vehicle Safety Act" President Lyndon B. Johnson noted that 29 American solders had died during the recent Labor Day weekend. During the same period 614 Americans had died in automobile accidents [193]. 1966 would have been at the height of the Vietnam War, which took place between 1st November 1955 and 30th April 1975.

6 In 2009, a number of safety incidents involving Mini vehicles build between 2001 and 2007 were reported. Although a power steering system failure was the cause, the vehicle still met its regulatory requirements so the manufacturer was not obliged to undertake a product recall in Europe. With the manufacturer making the statement that "as with all modern cars, power steering systems are an added support function that makes steering easier, especially at slow speeds, when parking for example. They are not an essential component of steering systems and cars can be driven perfectly safely without power steering assistance." [102].

value. This has led to the development of consumer information like the European New Car Assessment Programme (Euro NCAP) *5 Star* rating. Under this voluntary scheme a vehicle satisfying the minimum type-approval safety standards would receive a zero star rating, while a vehicle that achieves excellent crash protection and avoidance performance would receive the full *5 Stars* [52].

Despite the significant progress made in road safety, in 2022 the World Health Organisation (WHO) estimates that approximately 1.3 million people die each year [201] as a result of road traffic accidents and more than half of those are vulnerable road users; i.e. pedestrians, cyclists and motorcyclists. Additionally, between 20 and 50 million people sustain non-fatal injuries – with many of those causing permanent disabilities. The WHO suggests that, for most countries, the cost of road traffic accidents typically equates to 3% of a country's gross domestic product. In 2017 the WHO launched *Save LIVES*, a package of strategies aimed at halving the number of road deaths and injuries by 2020; covering aspects such as speed management, infrastructure design and improvement, vehicle safety standards and survival after a crash. In the *Save LIVES* foreword Dr. Etienne Krug indicates that to improve requires policy makers to overcome challenges, "particularly fatalism, the misconstrued notion that road traffic crashes are accidental and nothing can be done to prevent them" [60].

This view that road traffic accidents[7] and hence deaths are not inevitable is the long-standing view of the Swedish *Vision Zero* policy. Central to this policy is the belief that no fatality and serious injury statistic above zero is ethically defensible. The responsibility for accidents is typically apportioned to the individual. However, the exponents of a *Vision Zero* policy argue that the responsibility for zero deaths extends to both the road infrastructure and vehicle designers [46].

7 Many safety activists question the use of the word 'accident' in relation to road traffic incidents, suggesting that by referring to incidents as 'accidents' is accepting a level of inevitability and randomness – rather than acknowledging that many incidents are preventable [46].

Euro NCAP have embraced the notion of Vision Zero with their *Road Map 2025 – In Pursuit of Vision Zero* launched in 2017 [53]. The road map has a vehicle technology focus and as such

> Euro NCAP will challenge vehicle manufacturers to offer the best possible technology as standard in all segments and countries, protecting not only car occupants of all ages but also increasingly addressing the safety of other more vulnerable road users.

Presented as a time line, the road map documents vehicle primary (pre-crash), secondary (crash), and tertiary (post-crash) safety features considered as the key technology enablers in pursuit of Vision Zero. These ADAS and AD safety features include primary safety features, such as Driver Monitoring (2020) and vehicle to vehicle communications (2024), secondary safety features such as pedestrian and cyclist safety (2022) and Rescue (2020), and tertiary safety features such as child presence detection (2022) [53].

1.5 THIS RESEARCH'S CONTRIBUTION

The introduction of *active* and *tactical* vehicle safety features changes vehicle control. Vehicle control is now a task shared between the human and the automation. The nature of this shared control task and its potential impact on driving and vehicle control is discussed in Chapter 2. Of concern to the author and safety experts in the field (see Chapter 3), and the motivation for this research, is the potential impact that greater automation and shared control has on automotive functional safety. Referred to here as the *functional safety paradigm* (see Section 1.2.2), the automotive functional safety lifecycle implicitly assumes that the driver is integral to vehicle control, and as a consequence will remain situationally aware. In this context, it is appropriate to simply switch off a failed vehicle system knowing that the driver will re-adjust accordingly.

The potential impact of Automated Driving (AD) on the functional safety paradigm raises the question "How can the safety of AD be assured under

different levels of shared human-vehicle control?" To make progress in this regard, this thesis contributes the following:

- A review of vehicle safety, risk management and controllability literature, from the perspective of an automotive functional safety practitioner seeking to achieve AD safety assurance, has highlighted functional safety lifecycle tools, techniques and assumptions requiring redress (Parts I & II).
- The development of a vehicle model and behavioural competency taxonomy that aids the automotive functional safety practitioner to conceptualise and describe the nature of an AD vehicle feature's shared control (Chapter 8).
- The development of a concrete hazard analysis method that uses the vehicle model and behavioural competency taxonomy to analyse shared control hazard causes (Chapter 9), together with an accompanying safety case argument pattern for that (Chapter 10).
- Undertaking a series of case studies has established that the method presented exhibits the qualities necessary to be deemed a *proof-of-concept* (Chapter 12). However, further evaluation work is needed to establish the method's viability as a hazard analysis method for use by automotive functional safety practitioners working in a product development environment.

1.6 CONCEPTS AND MODELS

Throughout this thesis the terms *concept*, *conceptualise* and *model* are used extensively. To remove any ambiguity, this section describes the meaning I give to each term, and consequently the meaning that should be taken from these terms when encountered within the thesis.

1.6.1 *Concepts*

A *concept* is defined as an abstract idea [122] and typically describes a basal idea that underpins principles, thoughts or beliefs. Although philosophy research provides multiple interpretations of what should be understood by a concept [191], this research understands a *concept* to be a mental representation or visualisation of an idea, used to reason about (i.e. *conceptualise*) that idea. Consequently, the title of this thesis "Conceptualising automated driving shared control hazard causes" should be understood to mean the formation of an abstract representation of hazard causes that are attributable to AD shared control.

1.6.2 *Model*

A *model* can be thought of as a representation or copy of something physical, with the act of *modelling* being to create a representation or copy of a physical thing [124]. The reasons for creating a model are numerous, but typically one might wish to create a model to explore a particular property of a physical object or system, or to make predictions about future events. For example, a high-rise building scale-model could help evaluate the wind pressure experienced by the structure itself or by pedestrians walking around beneath [157], while a mathematical model might be used to make future statistical predictions about traffic accident rates [124].

Like the examples given above, the models described within this thesis have been created to explore some property of a system or to facilitate making future predictions about that system. Unlike the examples above, unless stated to the contrary, the models described herein represent concepts associated with the system rather than representing physical properties of the system itself.

Models of this type are often referred to as *mental models*. A mental model describes how an individual reasons about some system idea or principle [192]. For example, in the context of hazard analysis, that might be a

conceptual system model that aids the analyst to identify potential hazard causes (i.e., to make predictions about the likely future system behaviour). Depending on the context, a mental model might include a representation of the system environment and the relation between the system's constituent parts. In addition, the mental model will likely include an individual's intuitive understanding of cause and effect gained through practical experimentation with the system; as is typically the case for the car driver interacting with their vehicle's automation.

2

THE ART OF DRIVING

2.1 WHAT IS DRIVING?

Those of us who are qualified and have been driving for a period of time will likely take the task of driving for granted. We understand what it means to drive a car, but would probably find it difficult to articulate each control action and concept used. To explore the notion of driving and to inform how automation might change the assumptions on which the MISRA VCM is based, a review of driver behaviour modelling research was undertaken. Combined with a review of the literature pertaining to situational awareness, this helped inform how automation might affect driving and hence vehicle safety.

2.1.1 *Modelling the driver*

Driver behaviour can be thought of as resolving the multiple possible actions (or driving related subtasks) and their effects that are presented simultaneously in a dynamically changing environment [36]. Driver behaviour research can be traced back to Gibson and Crooks' 1938 research. However, it is suggested that driver behaviour research stagnated in the 1970's until in 1985 Michon published his seminal Hierarchical Control Model (HCM). The HCM (Figure 2) has three layers[1] (Strategy, Manoeuvring and Control), and Michon suggests

1 This research uses the terms Strategy, Manoeuvring and Control to described the three hierarchical control layers. These are also referred to by some researchers as Strategic, Tactical and Operational layers respectively. Abbink et al. suggests that a forth Executive Layer can also be added. The Executive Layer represents the neuromuscular control loops required to action Control Layer tasks [1]. This research groups Executive and Control tasks together within the Control Layer.

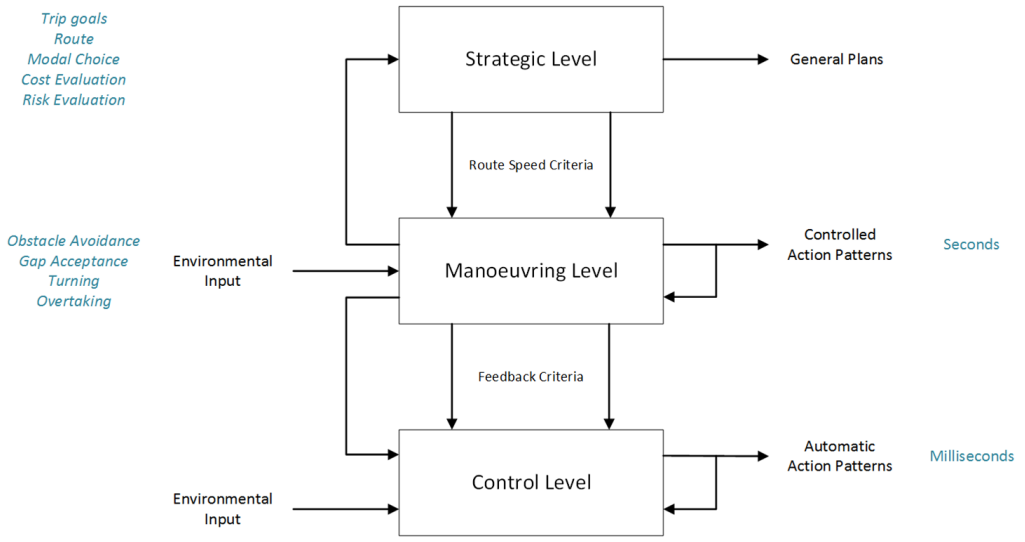


Figure 2: Michon's "hierarchical control model" (reproduced from [107])

that to be comprehensive a driver model should contain all 3 layers and provide information flow control between the layers [107]. The HCM has formed the foundation of subsequent driver conceptualisations [100], although contemporary driver models exist [28, 136, 188] their purpose has tended to mimic driver behaviour in a dynamic vehicle simulation context, rather than conceptualising driver behaviour in a hazard identification and analysis context.

2.1.2 *Situation awareness*

Fundamental to a driver's ability to safely control their vehicle, in a continually changing environment, is developing and maintaining an awareness of the objects and threats in that environment. The process of becoming and remaining "coupled to the dynamics of their operational environment" is termed Situation Awareness (SA) [145]. Salmon, Stanton and Young suggest that what SA actually is depends on the perspective from which SA is viewed [135]. Endsley's 3 Level Model views SA from the perspective of the individual. In this context SA exists as information within an individual's head or mind. In contrast, Ackerman's description of SA as applied in a military domain context,

views SA as "something that exists in the world" [145]. From this perspective, SA exists as information viewed on displays or other physical artefacts. The third perspective described by Salmon, Stanton and Young is a systems or human factors perspective. This perspective views SA as an emergent property of a socio-technical system, and is characterised by the distributed cognition that takes place between people and their environment [135]. In a socio-technical systems context Distributed Situation Awareness (DSA) [146] is the compelling notion.

Published in the 1980's, still popular and widely cited[2], Endsley's 3 Level Model [47] considers an individual's SA as:

- **Level 1 SA:** perception of the elements in the environment relating to the current task
- **Level 2 SA:** comprehension of the situation, involving comprehending data from Level 1 (i.e. the significance of objects and events), and
- **Level 3 SA:** the projection of future states, involving predicting the future states of the system and elements using a combination of Level 1 and 2 SA related mental models.

Endsley's Model has been used in combination with Michon's HCM to consider the SA needed by each hierarchical control level [84, 104], giving a concept in which the criticality of SA mental models [134] and factors affecting SA, such as distractions [153], might be considered in the context of vehicle safety. The ability to model both SA and driver awareness, and to represent those properties together within the Enhanced Vehicle Control Model (EVCM) (see Chapter 9 page 123), was the motivation for using Endsley's 3 Level Model for this research. This does mean that the EVCM reflects an individual's perspective of SA, rather than aligning with the systems perspective (e.g., DSA) that current research would advocate.

Although combining Endsley's model with Michon's HCM provides a compelling representation of driver cognition for this research, the 3 Level Model does have limitations that should not be overlooked. A criticism levelled at Endsley's

2 Having > 10,000 citations (checked January 2022).

Model by Salmon, Stanton and Young is that it assumes that SA is obtained and maintained by processing environmental inputs in a linear sequential way. However, not all driving tasks conform to this linear feedback control paradigm. Many tasks are feedforward in nature, and can be undertaken successfully with minimal input [145]. Many drivers will be familiar with the ability to arrive at ones destination, with no recollection of the precise drive to get there. This phenomenon demonstrates how experienced drivers can maintain SA without needing to perceive all elements in the environment. As Salmon, Stanton and Young note, viewed through the lens of Endsley's 3 Level Model, the "driving without attention" example would incorrectly imply poor quality driver SA [135].

While modelling SA from the perspective of the individual might be appropriate when considering shared control interactions between the driver and the automation in their vehicle, this perspective is less suited to more complex interactions. Particularly, as the different road users (e.g., pedestrians, cyclists, drivers) in a traffic system will likely have and need different levels of SA [135]. In addition, these differing SA levels combined with potentially conflicting goals, can result in "DSA breakdowns", and it is these DSA breakdowns that can themselves cause accidents [135]. To model these complex human machine interactions and DSA in a traffic infrastructure systems (e.g., multiple automated vehicles interacting, or automated vehicles interacting with other road users) a different model of SA is needed. A viable example of which is the Event Analysis of Systemic Team-work (EAST) method [185].

2.1.3 *Driving behavioural competencies*

Having made the observation that driving is a skill that soon becomes automatic, second-nature and potentially difficult to articulate, the need for a precise definition of what it means to *drive* became evident early on in this research – particularly while seeking to give precise meaning of each element of the EVCM diagram (see Section 8.2). To fit such a construct, this 'driving' definition would also need to embrace the different types of driving

task and align with Michon's Strategy, Manoeuvring and Control Levels (see Figure 2). Taxonomies describing the automation's capability requirements do exist [173, 200]. However, with the exception of Walker, Stanton and Salmon's Hierarchical Task Analysis of Driving (HTAoD) [186], a suitable taxonomy that describes 'driving' in its entirety was not forthcoming.

Consequently, to support this research, a driving competencies taxonomy has been developed (see Appendix A for the complete taxonomy), which draws extensively on the HTAoD [186]. The HTAoD defines four categories of driving task: "perform basic control tasks", "perform operational driving tasks", "perform tactical driving tasks" and "perform strategic driving tasks". Considering these categories together with Michon's control hierarchy it becomes evident that "perform basic control tasks" are sub-one second tasks that align with *control level* tasks. While "perform operational driving tasks" and "perform tactical driving tasks" are reflective of Michon's *manoeuvring level* tasks that might take a few seconds to perform. And "perform strategic driving tasks" perhaps unsurprisingly aligns to *strategic level* tasks in the HCM. The full HTAoD contains as many as six levels of abstraction for each category. As will be seen, for the hazard analysis undertaken by this research (see Chapter 8 and Chapter 9) using only two levels of abstraction has proven sufficient.

2.2 HOW DRIVING IS CHANGING

As discussed in Chapter 1 driving is a much safer endeavour today than it was 100 years ago. However, with more than a million people dying each year as a result of road traffic accidents and tens of millions more becoming permanently disabled [201], the global human and monetary cost each year is staggering and somewhat incomprehensible. As discussed in Section 1.4, such grim statistics are a clear motivation for National and International initiatives, such as *Save Lives* [60], *Vision Zero* [46] and Euro NCAP's *In Pursuit of Vision Zero* [53]. While initiatives of this type do address a breadth of road safety facets, vehicle based automation is clearly seen as a key enabler; particularly in a consumerism and brand differentiating automotive marketplace.

2.3 HIGHLY AUTOMATED DRIVING

2.3.1 *Automated systems taxonomy*

Terms such as "automated", "highly automated", "autonomous", "self-driving" and "autonomy" have become commonplace during the last decade – not just in the automotive industry, but in general conversation too. The ambiguity in the terms used, together with differing capability of AD vehicle features entering the marketplace, is probably the motivation behind many institutions seeking to categorise the different levels of automation – examples include, the German Federal Highway Institute [59], the National Highway Traffic Safety Administration (NHTSA) [5], and the Society of Automotive Engineers (SAE) [141]. Of those cited, it is the SAE's J3016 *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles* taxonomy that the industry has gravitated towards.

The SAE J3016 taxonomy seeks to provide a definition for the complete capability range of Automated Driving (AD) levels, thus it "can be used to describe the full range of driving automation features equipped on motor vehicles in a functionally consistent and coherent manner" [141]. Where an AD feature refers to a vehicle system that performs part or all of the Dynamic Driving Task (DDT)[3] for a sustained period of time; from completely manual driving with no automation at *Level 0*, through to fully automated driving under all conditions at *Level 5*. The levels within the taxonomy are represented pictorially in Figure 3 and the key characteristics of each level are described below.

3 SAE J3016 defines the DDT as "All of the real-time operational and tactical functions required to operate a vehicle in on-road traffic, excluding the strategic functions such as trip scheduling and selection of destinations and waypoints, and including without limitations: lateral vehicle motion control via steering (operational); longitudinal vehicle motion control via acceleration and deceleration (operational); monitoring the driving environment via object and event detection, recognition, classification, and response preparation (operational and tactical); object and event response execution (operational and tactical); maneuver planning (tactical); and enhancing conspicuity via lighting, signaling and gesturing, etc. (tactical)." [141].

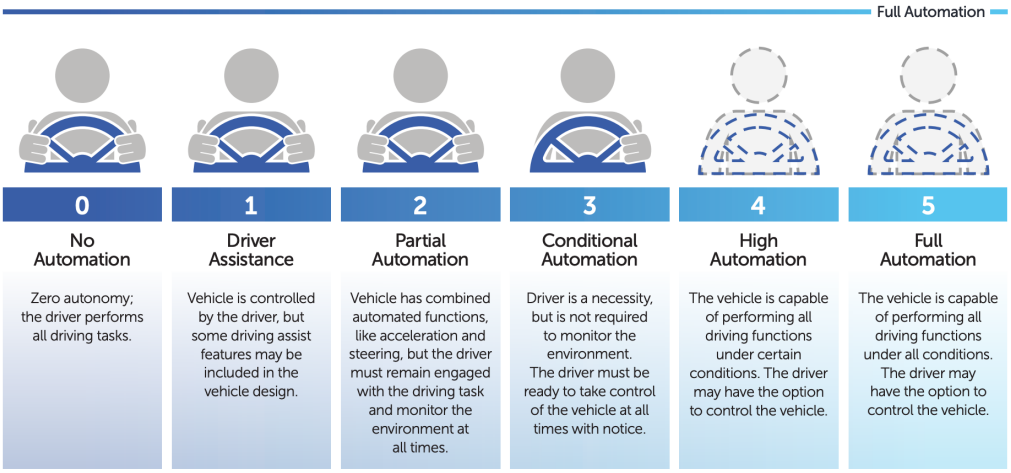


Figure 3: Pictorial representation of the SAE J3016 6 levels of automation (picture reproduced from [115])

In addition to the DDT, there are three further concepts within SAE J3016 worthy of mention: OEDR, DDT fall-back and ODD. The Object and Event Detection and Response (OEDR) is defined as a subtask of the DDT associated with monitoring the driving environment. It is perhaps analogous to the behavioural competency of *perform surveillance* (see Section A.1) although SAE J3016 refers to the competency in a more mechanistic way, that is "...(detecting, recognizing, and classifying objects and events and preparing to respond as needed) and executing an appropriate response to such objects and events." The DDT fall-back is the function that the user undertakes when the automation requires them to regain vehicle control. This could be because the automation has experienced a system failure or because environmental conditions / circumstances have changed causing the vehicle to no longer be operating within the Operational Design Domain (ODD) – for example, the vehicle is approaching a construction zone, or changing weather conditions (e.g. fog) adversely affecting a vehicle's vision system.

2.3.1.1 *SAE Level 0 – No driving automation*

SAE Level 0 represents full manual control. That is, the human driver undertakes all driving tasks all of the time. The vehicle may include supportive

safety systems, such as ABS or ESC, that can override the driver's inputs to help maintain vehicle stability, but the driver retains ultimate vehicle control.

2.3.1.2 *SAE Level 1 – Driver assistance*

SAE Level 1 *driver assistance* vehicle features typically automate one vehicle motion control task, with the driver being responsible for all remaining tasks. Vehicle longitudinal control is the driving task most commonly automated – with vehicle features like cruise control and Active Cruise Control (ACC) being common place examples. It is worth noting that even when a feature like ACC is operational, the longitudinal control task is still shared. This is because the driver may choose to override the automation at anytime. But in addition to that, there will be scenarios (such as another vehicle cutting in) where the automation may request the driver to undertake DDT fall-back and intervene to maintain vehicle safety. With such limits in authority and performance SAE J3016 defines OEDR and DDT fall-back as being the driver's responsibility. Although SAE J3016 describes the ODD as being limited, in reality few ADAS features will prevent the driver from activating the system at anytime. For example, if one chose to approach a roundabout or road junction with ACC active the system will do nothing to prevent such misuse.

2.3.1.3 *SAE Level 2 – Partial automation*

SAE Level 2 *partial automation* automates two vehicle control tasks, so is the natural progression from Level 1. Although the automation is expected to undertake lateral and longitudinal vehicle control for sustained periods, while inside the ODD, the driver is still expected to undertake all OEDR tasks. In addition, the driver must remain fully alert as they may be called upon at any moment to fulfil the DDT fall-back function. Like Level 1, both lateral and longitudinal vehicle control are potentially shared between the driver and the automation in certain driving scenarios.

2.3.1.4 *SAE Level 3 – Conditional driving automation*

SAE Level 3 *conditional driving* automation is perhaps the most controversial of the defined levels [18, 138]. This is due to the 'paradigm-shift' that occurs between Level 2 and Level 3 regarding driver and automation responsibility. Pictorial representations of the SAE taxonomy typically make a graphical distinction between SAE Levels 2 and 3 (Figure 3 is a case in point). The automation levels discussed until now have all defined the driver as being responsible for the OEDR tasks while the automation is active. At Level 3, for the first time, *conditional driving* automation is expected to undertake all OEDR tasks while the automation is operational. However, the driver is still expected to be DDT fall-back ready. Although, interestingly in this context the human is typically referred to as the "fall-back ready *user*" rather than the "fall-back ready *driver*". For Level 3 automation the ODD will still be limited, and with the additional OEDR responsibility, the system will almost certainly include functionality to detect ODD limits and to ensure that the driver has DDT control before the vehicle exits the ODD.

2.3.1.5 *SAE Level 4 – High driving automation*

Although the ODD is still limited at this level, at SAE Level 4 the automation is responsible for the full DDT, all OEDR tasks and for DDT fall-back while the vehicle is inside the ODD. With the vehicle occupants never needing to undertake any part of the DDT it is conceivable that a vehicle incorporating Level 4 *high driving* automation will be built without driver controls. Applications deploying Level 4 automation might include driverless taxis or public transport shuttles operating from A to B in geofenced[4] areas. Typically, vehicles of this type will still require an amount of human control, with a remote operator often being needed to position the vehicle at the beginning of the day or as part of service and maintenance operations.

4 "Geofence" refers to a "virtual geographic boundary, defined by Global Positioning System (GPS) or Radio Frequency Identification (RFID) technology, that enables software to trigger a response when a mobile device enters or leaves a particular area". It can also be used as a verb to describe the act of creating a virtual geographic boundary [123].

2.3.1.6 *SAE Level 5 – Full driving automation*

Completing the SAE taxonomy is SAE Level 5 *full driving automation*. Like Level 4, the automation is responsible for the entire DDT, all the OEDR tasks and for DDT fall-back, so like Level 4 a user is not expected to intervene in the vehicle's operation. As a consequence, the only difference between Level 4 and 5 automation is the ODD. For Level 5 it is unlimited. At the time of writing no full driving automation production systems exist.

2.3.2 *AD feature evolution*

Advanced Driver Assistance System (ADAS) is the umbrella term for a range of vehicle features designed to aid, warn and assist the driver to undertake simple tasks related to the DDT. Although the term ADAS has only become commonplace in recent years, early example of ADAS vehicle features can be traced back as far as the 1950's. In 1959 Cadillac introduced its Cyclone concept vehicle that included a RADAR based collision avoidance system [38]. Some common ADAS features, with their first introduction dates, are summarised below [97, 111]:

- **Driver Aids:** night vision (2000), adaptive front headlights (2006), front and rear cameras, and surround vision systems (2007), traffic sign recognition.
- **Driver Warnings:** park assist (2002), forward collision warning (2003), lane departure warning (2005), blind spot monitoring (2006), rear cross traffic (2006) and driving monitoring (2006)
- **Driver Assistance:** self-parking (2006), adaptive cruise control (2007), forward crash mitigation (2008), lane keep assist (2010), and pedestrian avoidance (2014)

In recent years these basal vehicle features have been combined to develop partial automation systems. Two examples are Traffic Jam Assist (TJA) and Highway Assist System (HAS). TJA can undertake longitudinal and lateral

vehicle control at low-speed in dense traffic situations. HAS again undertakes longitudinal and lateral vehicle control, but at highway speeds, with some HAS also supporting the driving during overtaking manoeuvres.

Within this thesis ADAS, partial, and conditional automation systems are used to explain and evaluate the concepts and contributions presented. Use case descriptions of the vehicle features used are included in Appendix B.

Although designed to aid the driver, reduce the likelihood of human error and thus accidents, multiple research studies have highlighted potential pitfalls associated with vehicle automation. Although its focus was the process industry, Bainbridge's seminal paper exploring the "Ironies of automation" [17] set the tone for the human factors automotive specific research that has followed. Research projects such as CityMobil (European), OPTIVE (Sweden) and Easy (UK) have investigated driver behaviour in relation to ADAS. Specifically, potential human factors issues [97], driver resumption of DDT control performance [105], and the influence that the automation's level and type might have on driver attentiveness and engagement in secondary tasks [30]. While studies such as FORWARN (UK) have sought to investigate the impact of distraction on driver performance [90] and BAMADAS (Netherlands) considered how automation concept improvements might improve driver engagement [56].

Irrespective of the substantial funding and research effort that has been focused towards improving sharing the driving task between the human driver and the automation, high profile accidents and examples of potential driver misplaced trust and miss-use continue to appear in the media [62, 159] and on YouTube [190]. And as Undercoffler observes, "While these systems are touted as "hands-free," they are not "attention-free." They require the driver to be paying attention at all times, even if their hands aren't on the wheel" [176]. Changing the perception of what it means to *automate* the driving task, in comparison to other commonplace endeavours, is perhaps the key enabler to the safe introduction of greater automation. As Stanton eloquently described during a keynote address, the interaction that humans are expected to have with vehicle automation is atypical. For example, when placing a load of dirty cloths into

the washing machine, adding the detergent, selecting the wash programme and then presses 'start', one doesn't then "pull up a chair in front of the washing machine and watch it complete the task" [148].

2.3.3 *Shared control*

Abbink et al. suggest that the ability for machines to be "fully autonomous always and everywhere is a myth" [1]. So, with the exception of this fully autonomous utopia or purely manual control, there is the need for some sort of communication[5] or sharing of control to take place between the human and the automation. An inevitable consequence of system control success relying on two 'intelligent agents' is complexity, with a commonly held view being that it is only the purely manual or purely automated use case that will be completely immune from conflicts in control [203].

Like the interaction themselves, the definitions and language surrounding human-machine communication and control appears no less complex. As the capability of automation increases, three terms emerge describing the human-machine relationship: Human Machine Interaction (HMI), Human Machine Collaboration (HMC), and more recently Human Machine Teaming (HMT). With HMI being reflective of a start-stop button press style relationship that sees the human firmly in command [154]. In industrial setting, collaborative style (HMC) relationships exist where industrial robots undertake the repetitive tasks previously carried out by human operators, with the human still potentially intervening and providing oversight [91]. Still largely a research pursuit, HMT is reflective of a human-machine relationship where the human and machine agent's aligned goals are worked on together, in a more dynamic and less predicable way [187].

Although referred to in the context of HMT above, machine goal setting is not a new concept. In their review of shared control systems Abbink et al. discuss a decades old term, *supervisory control*, in which a machine (without

5 Interestingly, *communication* originates from the Latin *communicat-* meaning *shared*, which is from the verb *communicare* [121].

being fully autonomous) can be set a goal to complete unsupervised. However, one challenge identified with such interactions is establishing the method of communication between the human and the machine [1]. This has lead to the emergence of two further terms: *shared control* and *traded control* [188]. During *shared control* the human and machine agents are active together, while *traded control* sees the agents taking turns being active. Abbink et al. note that while both traded and shared control can be applied to the task hierarchy, it is the Strategic Level and Manoeuvring Level tasks that tend to be traded, with true shared control occurring at the Control Level. In an automotive context, the driver enabling, overriding and cancelling ACC would be an example of traded longitudinal control. While LKA is an example of *haptic* lateral shared control and steer-by-wire an example of *input-mixing* lateral shared control [1].

While introducing their "hierarchical framework of shared control" Abbink et al. highlight the importance of considering all potential HMI combinations during the design and evaluation of shared control; particularly in scenarios where agents are learning from one another [1]. Comparable to the behavioural competency responsibility matrix introduced in Section 8.4, Abbink et al. identify the type of interaction (i.e. performed independently by an agent, traded, or shared) taking place at each level in the task hierarchy (i.e. Strategy, Tactical, Operational, Evaluation [STOE]). They suggest that by embellishing the interaction definition with the type of behaviour required from each agent (i.e. Knowledge, Rule or Skill [KRS]) system capability requirements (e.g. the need for mental models) can be better understood. Although for hazard analysis this research uses a less-granular HMI definition (see Section 8.4), incorporating the STOE-KRS framework into an AD feature's item definition could be a beneficial enrichment[6].

Having introduced some of the terminology used in the field of human-machine communication and control, it should be noted that this research makes no attempt to apply these terms precisely. This is particularly true of *shared control*, which by definition excludes traded tasks, because the definition requires both agents to be "temporally congruent" [1]. With the research

6 The potential benefit of incorporating the behavioural competency taxonomy responsibility matrix into an AD feature Item Definition is discussed in Section 12.1, page 168.

fields of human-machine communications and shared control being too large to cover appropriately here, I uses the term *shared control* in its broadest sense, to describe situations where some level of human-machine interaction is taking place in pursuit of task completion – be that specifying which agent has responsibility for doing, monitoring, or achieving safety in the context of the behavioural competency taxonomy (see Table 4 page 128), or be that reasoning about the nature of hazard causes when addressing the loss scenario questions (see Figure 17 page 137).

2.3.4 *Technologies deployed*

Various frameworks have been used to describe the decision making process (by both humans and by automation), such as OODA Loop [194], SPA [143] and SUDA [27]. The SPA and SUDA frameworks are typically used by design engineers to describe the primitives that automated systems must implement, namely: *sense, plan, act* (i.e. Sense Plan Act (SPA)). Arguably the ability to *sense*, but also to correctly *understand* (i.e. Sense Understand Decide Act (SUDA)) what is being sensed, is an important fact in a system safety context.

If technology is to replace the human in Figure 1 (page 19) and undertake *Driver Control*, then the automation must be capable of the following:

- **sense:** to sense the characteristics of the environment in which the vehicle is operating. This includes the identification of obstacles and navigational aids. In an automotive context this includes sensing other road users (obstacles) and road markings (navigational aids).
- **understand:** the ability to build a 'picture' of the vehicle's operating environment. By using temporal sensing data and data fusion (see Section 2.3.4.1) the system must build an accurate model of its environment.
- **decide:** based on its environmental perception, the system must then decide what to do next. In an automotive context this involves using the vehicle's 'localisation' knowledge of its current and future position, with

a prediction of the path and movement of other road users, to plot the safe future trajectory for the vehicle.

- **act:** the ability to control the vehicle's lateral and longitudinal movement. This typically involves modulating the vehicle's steering, braking and propulsion actuators.

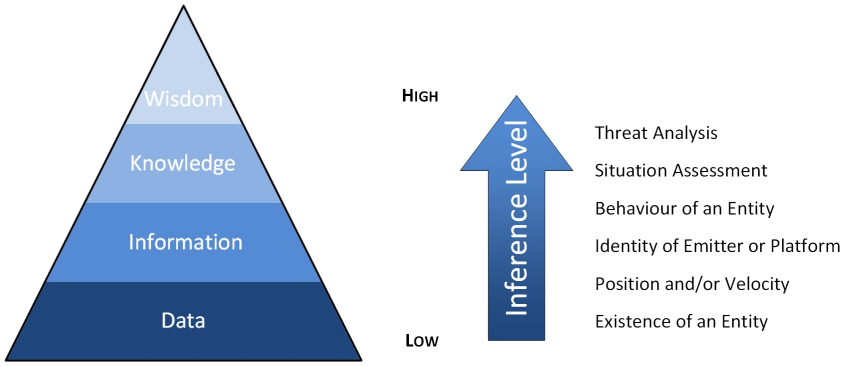
2.3.4.1 *Data fusion*

In an article describing the updates being made to Tesla's Version 8 Autopilot [167] the need to 'fuse' data from different sensor types is evident. For example, in a world seen by RADAR people are invisible, an aluminium drinks can's concave base can lead to its size appearing magnified, and an overhead metal gantry being approached downhill many appear as a solid object across the road. With all sensor technologies currently available having limitations [199], the outputs from an array of sensor technologies is typically used to reliably sense the environment under all operating conditions. Therefore, "the aim of a data fusion process is to maximize the useful information content acquired by heterogeneous sources in order to infer relevant situations and events related to the observed environment" [103]. Data fusion supports reliable sensing, but to *understand* the behaviour of artefacts in the environment requires *a priori* knowledge to then correctly interpret what is being sensed.

2.3.4.2 *Data inference*

Inference in an automation context involves taking data from a number of low level sources, and then inferring information from that data at a higher level in the information hierarchy. For example, inferring vehicle speed from individual wheel velocity measurements, or inferring a target object's classification by comparing the speed of that object with the vehicle's current speed. Data-Information-Knowledge-Wisdom pyramids [16] provide a useful way to visualise such an Inference Hierarchy (see Figure 4) [68]. Hall and Llinas suggest that by (correctly) combining the data from multiple sensors a more accurate determination of a moving objects distance and direction can

be obtained than could have been achieved from using data from just one sensor [68]. However, they do advise caution, noting that care is needed during the design process to avoid situations where the system becomes infeasible to implement and achieves worse results than would have been possible from one carefully chosen single sensor. As highlighted in [108], the inference hierarchy helps visualise the difference in perception that may exist between the automation and the human driver – a potential important consideration for ADAS, partial and conditional automation safety, where vehicle control will be shared or passed between the human driver and the automation.



(a) Hierarchy pyramid (from [16]) (b) Inference hierarchy (from [68])

Figure 4: Data fusion inference hierarchies

2.3.4.3 *Object detection*

The topic of object detection is too broad to do the subject justice here. At a superficial level, to support the *understand* primitive discussed in Section 2.3.4, the object detection function must correctly identify and classify objects in the vehicle environment. This is both the identification and classification of objects that the vehicle should avoid to maintain vehicle safety, and the identification and classification of objects (e.g. pedestrians, cyclists) to which the vehicle could inflict harm. It is in support of object detection where automotive systems have seen the introduction of neural networks and machine learning algorithms; which because of their complex and opaque nature their introduction challenges the established functional safety paradigm (see Section 1.2.2) yet further.

Typically object detection involves a training phase, followed by the detection phase. The detection phase occurs at 'run-time' and typically incorporates two further steps: feature extraction or *hypothesis generation* and object classification or *hypothesis verification*. This two step approach is taken to reduce the overall computation effort needed to achieve object recognition at run-time. Hypothesis generation involves determining where within the image objects of interest might reside. For example, to identify parked cars within an image the system needs to know where in the image one will typically find parked cars. Then once candidate objects are identified, further algorithms may then use other image artefacts to increase object detection confidence. For example, using areas of contrasting shadow, or the 'sharp' edges of a windscreen to increase the certainty that the object detected really is a parked car. To identify moving objects in the scene the concept of *optical flow* is typically used. This involves detecting the relative motion of each pixel in a sequence of time series images to determine the image flow (Figure 5 visualises the *optical flow* overlay for a forward facing camera). The optical flow overlay for an approaching vehicle would see the pixels diverging, while for a departing or overtaking vehicle the optical flow overlay would be seen to converge. Stationary objects will also have optical flow relative to the target vehicle - providing the target vehicle is moving of course, which does highlight a limitations of this type of technique [8, 156].

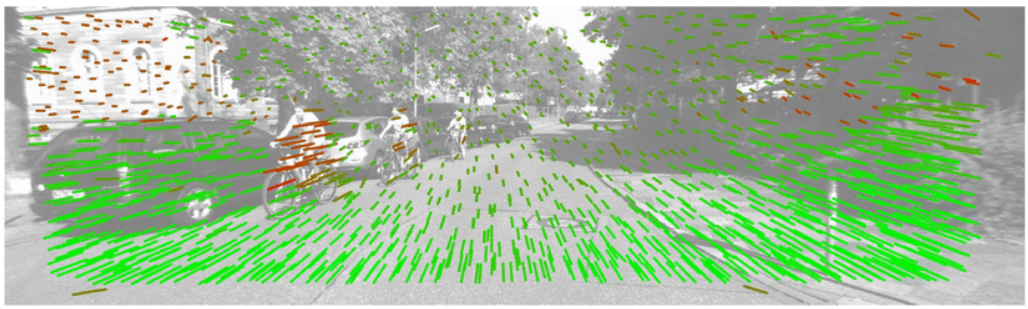


Figure 5: Example forward facing camera footage with *optical flow* overlay (from [73])

Once hypothesis generation has identified potential objects within the scene, hypothesis verification then seeks to validate the classification of those objects. Template and appearance based validation are two methods typically deployed. The former method looks for a particular pattern within an image, while the

latter method uses the difference between image classes to 'teach' the system about a particular object type. So for example, a template based validation might use the 'U' shape pattern made by a vehicle's windscreen as validation. In contrast, teaching the system using appearance-based validation might use two sets of images, with one set of images having vehicles present, while the other set of images does not. Appearance based hypothesis verification is a more reliable method than template based validation [8], but successfully training a system with either method is a delicate balance – both in terms of the costs involved in providing sufficient training examples, while at the same time avoiding over-training pitfalls like *overfitting* [144].

3

RESEARCH AIMS

3.1 CHALLENGES OF HIGHLY AUTOMATED DRIVING

It was important for this research to engage with the key automotive stakeholders to understand their needs. Thus maximising the potential for this research effort to have a tangible and positive benefit in the automotive domain.

An invitation to run a workshop on AD safety, at the IQPC Functional Safety Conference in March 2017, presented an unmissable opportunity. As it provided the forum in which to engage with other automotive functional safety experts and to undertake structured qualitative research probing what those experts considered the contemporary AD safety engineering challenges to be [110]. Forty one functional safety experts, practitioners and development engineers took part, predominately from automotive suppliers and academia. The workshop began by providing the participants with some background material, by discussing the notion of controllability, the implicit assumptions made by ISO 26262, and the level of control relinquished by the driver with increasing automation [141]. In addition, the introduction briefly discussed two pivotal research papers by Bainbridge and Parasuraman and Riley relating to human interaction with automation [17, 128].

The 'World Cafe' style [168] workshop was approximately 4 hours in duration. To make best use of that time, the delegates were split into 5 groups to consider the following topics: ISO 26262's current scope, the hazard analysis and risk assessment process, the notion of controllability, and safety concepts. Using a Thematic Analysis technique uncovered a number of themes emerging from the delegates' transcripts. The themes relating to understanding AD risk and safety are listed in Table 1, while the themes relating to interactive complexity

Table 1: A new risk model

UNDERSTANDING RISK AND SAFETY: Understanding the risks associated with, and demonstrating the safety of, automated systems in the context of today's safety lifecycle will be challenging	
Theme 6	*Controllability and Automated Systems:* Today's notion of controllability cannot be applied without change and is further affected by the level of automation.
Theme 7	*Expanding Notion of Controllability:* The notion of controllability needs to be expanded to consider controllability for the driver, for other road users and for the system itself.
Theme 14	*Multi-Level HARA:* Multi-level HARAs are required for automated systems that cover the broad scope of the vehicle, the driver, and the environment.
Theme 15	*Complexity Impacts HARA Effectiveness:* Some aspects of the HARA will not change, however the effectiveness of such analysis will be influenced by more complex automated systems.

appear in Table 2. Themes of direct relevance to this research are: the need for the hazard analysis and risk assessment process to have a broader scope, the challenge of managing increased complexity, and potentially modifying the notion of controllability.

3.1.1 *Hazard Analysis and Risk Assessment*

The topic of automotive Hazard Analysis and Risk Assessment (HARA) is discussed further in Chapter 5. For the workshop participants a number of potential issues emerged (see Table 1), with the key 'take-away' being the need to broaden the scope of hazard analyses.

The participants noted that although "HARA is focused on 'malfunctioning behaviour'", a HARA could potentially support the analysis and correct specification of the intended function. However as one participant put it "the specification of the correct function cannot contain a complete description of the world."

Many factors have the potential to influence the number of operational situations that should be considered during the hazard analysis process. One factor identified was the need to broaden the scope of the analysis beyond a single vehicle: "HARA analysis of [the] item would need to be done on vehicle or even road infrastructure level". Another factor is the potential for either the driver or the system to be in control. Consequently the participants view was that the hazard analysis should consider both circumstances.

As discussed further in Chapter 5 and Chapter 6 the notion of controllability is used by automotive safety practitioners to classify accident risk; by considering what the likelihood is that the driver will be able to control the hazard and thus avoid an accident. Even though it may be appropriate to remove controllability from the risk model for delegated[1] driving it's influence is still relevant while the driver continues to share the driving task with the system. However, for shared driving participants suggested that "controllability needs to be thought of differently", as "situations leading up to the hazardous situation" may have an influence on controllability. Additionally, working with automation may affect the driver's behaviour and awareness of their surroundings [106], as a consequence the "time to get driver back in the loop" may be longer.

Intuitively one might conclude that greater automation will drive complexity into the HARA process and increase its scope. This was a view shared by the workshop participants. The participants felt that the scope of the analysis should be increased to consider the effects of the intended function as well as the malfunctioning behaviour of the system. Although in principle this does seem an appropriate step to take, as one participant postulated "what does it mean for an autonomous system to fail?"

Feedback from the participants suggested that a hierarchical HARA would also result in a *"sphere of influence"* challenge. For example, a system manufacturer might carry out a system level HARA making assumptions about the vehicle into which the system would be fitted, while the vehicle manufacturer might carry out a vehicle level HARA making assumptions about the infrastructure. Finally the infrastructure supplier might undertake a HARA for the infrastructure

1 For a discussion on shared and delegated driving please see Section 3.1.3.

equipment making assumptions about the influence the infrastructure may have over vehicle behaviour. It would then be reasonable to expect the safety case for such a system to reconcile those assumptions and to present a coherent final safety case argument. Although such an activity is not typically undertaken by the automotive industry, this type of approach does have precedence in other industries, such as aerospace.

Participants contemplated whether this idea of shared HARA responsibility still works in situations where different vehicle types are interacting with each other and the environment? For example, a number of different manufacturers' vehicles cooperating in a road-train? Perhaps in this scenario the road-train *"rules"* would need to be the subject of HARA analysis too? This idea of analysing the *"rules"* has been used on infrastructure projects such as the M42 Smart Motorway [9].

The participants observed that sheer scale of the HARA will also likely increase. That is, the risks associated with any given hazard may need to be considered for a number of hazardous events to take account of the different people in harms way. For example, one might conclude that a hazard outcome associated with a vehicle operating automatically might be for the vehicle to decelerate unexpectedly. Today's HARA would typically consider hazardous events in the context of the vehicle driver. However if the vehicle's brake system were under automatic control, then the analysis ought to consider the influence on other road users travelling behind. The level of connectivity between vehicles may further complicate the HARA for such a system, and the level of connectivity achieved could reasonably be expected to change from scenario to scenario.

In future hazard identification becomes a multi-dimensional problem, with the need to consider hazards resulting from interactions between the vehicle and the driver, and between vehicles. Safety concepts will need to address these new hazard causes, and not be limited to the detection and mitigation of failures within the vehicle control system. This added interactive complexity complicates the validation process, as does the assurance challenges associated

with systems that may learn, and whose behaviour may be adversely affected by the environment.

3.1.2 *Increased complexity*

A very clear message resonating through the data, which is applicable to this research, is increasing complexity (see Table 2). Architectures needed for automated vehicle operation clearly become more complex. No longer can the car be thought of as a *"satellite"* system, given that there is now a need to interface directly with one or more of *"The Cloud"*, other vehicles and roadside infrastructure. As a system of systems both the design and validation phases of development become more complicated than for a vehicle with little or no automation.

The workshop participants felt it appropriate for the public expectation to be that the risk of being involved in a road accident will greatly reduce with increased automation; particularly as in all likelihood the car owner will have paid a price premium for such vehicle features. This does however lead to a potential dichotomy between public perception and the real hazard landscape, given that these new technologies are complex and have the potential to challenge current design and development methodologies. This is particularly true of ISO 26262 that, although published in 2011, is derived from principles developed in the 1990s [83], and at that time it was observed that vehicle systems where "rather well defined, owned by [the] OEM and limited in scope, [and] self-contained." As a consequence one participant suggested that ISO 26262's "current scope/focus is/was for 'simple' systems" which means that "[ISO 26262] doesn't cater for [the development of] complex/autonomous systems."

Table 2: Interactive complexity

INTERACTIVE COMPLEXITY: Complex functionality, non-determinism and system of system interactions challenge the definition, analysis, verification and validation of automated systems; making the demonstration of safety difficult.	
Theme 1	*Adaptive & Learning Technology:* Demonstrating safety is challenging when the system includes adaptive or learning algorithms.
Theme 3	*Complex Interactions:* Complex interactions with the infrastructure and other systems (including legacy systems) make the definition of system boundaries and integrity assumptions difficult.
Theme 4	*Complexity of V&V:* Verifying and validating automated systems is challenging given the number of use cases and the complexity of the test environment.
Theme 5	*Structured Verification & Validation:* A structured approach to automated system verification and validation is required; incorporating layers of verification and different verification methods.
Theme 8	*Transition of Control:* Automated systems add complexity and introduce uncertainty into the transition of control from the system to the driver and vice versa.
Theme 9	*Vehicle Level Safety Concepts:* In order to support greater automation vehicle level safety concepts will be needed covering interfaces to the infrastructure and to other vehicles with varying automation levels.
Theme 12	*Safety Assurance:* A system of systems approach is needed to design and safety assure complex automated systems.
Theme 14	*Multi-Level HARA:* Multi-level HARAs are required for automated systems that cover the broad scope of the vehicle, the driver, and the environment.
Theme 15	*Complexity Impacts HARA Effectiveness:* Some aspects of the HARA will not change, however the effectiveness of such analysis will be influenced by more complex automated systems.

3.1.3 *The notion of controllability*

One workshop round explored the notion of controllability by asking the participants to consider the MISRA VCM (see Figure 1, page 19). Participants felt that intuitively controllability is "massively impacted by greater autonomy." Without automation, controllability can be viewed as solely the driver's ability to maintain control of the vehicle following a hazardous event. However, with increased automation the participants felt that controllability should be considered in the broader context of the driver's cognitive / emotional state, other road users, the level of automation involved, the vehicle's operating environment (e.g. road conditions, road layouts, driving cultures), and which system / function has failed. The fact that controllability was no longer solely about the driver was discussed.

The increased complexity of the transition between the driver and the system was acknowledged; with the potentially bidirectional nature of the driver-system interaction being noted. The transition from system to driver was widely acknowledged as being critical to safety, with many participants expressing the view that once the human has relinquished control the system should not ask them to retake control; particularly when safety relies on the driver making a correct and timely response. Who has ultimate control of the vehicle was discussed, and parallels drawn to what is referred to by the aerospace community as *mode confusion* [205].

Once out of the control loop the driver's behaviour also changes [93, 99]. Seppelt and Trent suggest that rather than having multiple automation levels, there should be just two types of automation [138] – *shared* and *delegated* driving. Shared driving is analogous to the SAE's driver assistance and partial automation levels (see SAE Levels 1 and 2 in Section 2.3.1, page 31) [141] where the responsibility for vehicle safety remains with the driver, and the system assists the driver in carrying out their normal driving control task. Seppelt and Trent reject the legitimacy of SAE Level 3 conditional automation, and instead align delegated driving with high driving automation (SAE Levels 4) and fully automated driving automation (SAE Level 5) – where the driver delegates the

whole driving task, and therefore is not expected to regain control to maintain safety. The comments received from participants supported this view.

3.2 WHAT THIS RESEARCH SEEKS TO ACHIEVE

Motivated by an interest in the notion of controllability in the context of AD and sharing many of the workshops delegates' "worries", this research seeks to develop the HARA process for AD systems.

When undertaking the HARA for a stand-alone vehicle feature (e.g. engine control, brake stability control) the MISRA VCM (see Figure 1, page 19) is a useful *thought experiment* aid. This is because the MISRA VCM allows one to conceptualise the impact that a vehicle system malfunction might have on vehicle behaviour, and hence the driver's ability to retain control in different driving contexts. Unfortunately, AD vehicle features break the model, by splitting the *Driver Control* bubble between the driver and the automation. In addition, the potential breadth and depth of the hazard analysis for an AD vehicle feature also poses a problem; both in terms of the levels of abstraction (e.g. the system, vehicle, and infrastructure levels discussed above) requiring consideration and the inherent complexity that dealing with those levels brings.

Consequently, this research seeks to answer the following question:

> How can the safety of AD be assured under different levels of shared human-vehicle control?

The research addresses the above question by contributing a hazard analysis method for shared control, comprising an Enhanced Vehicle Control Model (EVCM) (see Chapter 8), an analysis method that incorporates joint-cognition principles (see Chapter 9), and a safety case argument pattern for shared control (see Chapter 10).

Then to be deemed *"effective"*, the evaluation (see Part IV) seeks to ascertain to what extent the above research products exhibit the following properties:

1. The EVCM supports the conceptual modelling of AD vehicle features having shared control. And as a conceptual model the EVCM can be used to facilitate AD hazard analysis *thought experiments*.

2. When used together, the EVCM and behavioural competency taxonomy supports AD hazard analysis, by providing a conceptual vehicle control model for AD vehicle features that include shared control.

3. Focusing on behavioural competencies with shared actor responsibility and using cognitive principles to enhance *Classic* STPA loss scenario types, allows the EVCM and accompanying *Shared Control* Systems-Theoretic Process Analysis (STPA) method to identify hazard causes and hazardous situations associated with shared control that might otherwise remain undiscovered.

4. The supporting safety case argument methodology highlights potential loss scenarios attributed to shared control and emphasises how AD feature design modifications mitigate such hazard causes.

Part II

LITERATURE REVIEW

Chapter 3 posed the question "How can the safety of AD be assured under different levels of shared human-vehicle control?" A question that this research seeks to address by developing a conceptual model of shared control, together with an accompanying hazard analysis method.

Hazard analysis processes support safety assurance by identifying potential sources of harm (i.e., hazards) and classifying the risk associated with the hazards identified. This epistemic knowledge then typically informs engineering design decisions regarding the choice, integrity and prioritisation of risk reduction measures. Speaking from first-hand experience, automotive design engineers are most comfortable working with quantifiable absolutes. However, as revealed by this literature review, the decision-making process for the automotive safety practitioner is littered with subjectivity and uncertainty.

To provide context for the development, introduction and use of AD vehicle features, this literature review begins by discussing the commonly used, but arguably overloaded, terms of *risk* and *uncertainty*. It then looks at how industry has sought to model, analyse and manage risk, before finally looking at the notion of controllability. A concept that is fundamental to functional safety hazard classification and risk reduction, but that the introduction of greater automation certainly challenges, and maybe even breaks.

4

RISK AND UNCERTAINTY

'*Risk*' is a word used frequently[1] in everyday language. However, the meaning one takes from the word *risk* differs depending on the speaker, their discipline and the context in which the word is used. The car driver might consider overtaking on the brow of a hill too risky, an insurance actuary may calculate risk for the young driver who has recently passed their driving test, or a surgeon might discuss the risks (and benefits) of elective surgery with their patient. Given that the notion of risk is not unique to any one discipline, and can be either subjective or objective, differences in its derivation, interpretation and meaning are inevitable.

As introduced in Section 1.2.2, functional safety is defined as the absence of *unreasonable risk*. Given that this research seeks to contribute positively to AD functional safety, how one might interpret *risk* warrants further discussion, both from a theoretical perspective and from the perspective of the automotive practitioner seeking to conform with industrial standards. To support such a discussion, this chapter explores risk research literature, while Chapter 5 discusses risk from the perspective of engineering practice.

Relative to their early counterparts, modern passenger vehicles are undeniably safer. However, as passenger vehicle users we accept a level of risk each time we make a car journey, for the utility and convenience it brings. AD technologies increase vehicle capabilities, which presents the automotive consumer with new opportunities and capability expectations. However, the enabling technologies, like machine learning (see Section 2.3.4.3), that make AD commercially viable are themselves complex. A motivation for introducing AD technologies into vehicles is to positively influence vehicle occupant safety, by reducing the

1 Part I of this thesis is a case in point, with the word *risk* appearing 17 times in the introductory text.

occupants' risk exposure. However, typically these technologies are complex and not transparent, which introduces uncertainty.

This chapter explores the many facets of *risk* (and *uncertainty*). From the origins of the word itself, to the ways in which society perceives risk. Risk perception and uncertainty are of particular relevance to the AD discussion, and the chapter concludes the discussion on risk by considering risk perception and uncertainty in the context of technological change.

4.1 THE NOTION OF RISK

The origins of the word *risk* are discussed by Althaus in a interdisciplinary literature survey *"A disciplinary perspective on the epistemological status of risk"* [7]. Although it would appear that some disagreement exists as to exact origins of '*risk*' in natural language, it is suggested that by the 15th century the European sea going nations were using *risk* to describe either: the perils of sailing in uncharted waters or of the voyage itself, the uncertainties of sailing near cliffs or rocks, or the potential damage to valuable cargo.

Althaus summaries risk as "an ordered application of knowledge to the unknown" and suggests that each discipline "has a particular knowledge approach with which they confront the unknown and thus understand risk" and that each discipline has its own methodology for "enquiring and applying knowledge to uncertainty" [7]. As highlighted by the workshop delegates (see Section 3.1.1), one impact of AD on automotive system safety is the potential for complex or incomplete[2] prior knowledge, to challenge the current automotive HARA process and to increases future risk projection uncertainty.

Oxford Languages on-line dictionary [125] contemporary definition of risk provides two usage perspectives:

- ***noun:*** a situation involving exposure to danger
- ***verb:*** expose (something or someone valued) to danger, harm or loss

2 For example, if a new technology is being deployed for the first time.

In addition to being a noun or a verb, when used in natural language *risk* can have both positive and negative connotations. Some decision making disciplines "do treat risk as a neutral term" [133], and in some context risk is considered a positive[3], but it is the negative slant that has greater prevalence in natural language today.

Depending on the context, risk can be measured (quantitative risk) or it can be subjective (qualitative risk). Rohrmann suggests that science disciplines typically consider quantitative risk – with definitions focused towards probabilities and the likelihood of negative outcomes. Rohrmann contrasts this with social science, where he suggests that "the 'meaning' of risk is the key issue", and the qualitative aspects of risk are important. However, 25 years' experience in the automotive industry suggests to me that such a straightforward distinction between the natural and social sciences is misplaced; with automotive safety standards and guidelines discussing the elements that comprise the risk model in depth, and stressing the importance of documenting a reasoned rationale with any risk classification made [113, 162]. Indeed, this subjective "value judgement" is inherent in any risk identification technique [10].

The philosopher Thompson has discussed the nature of risk, the meaning conveyed by risk statements (in both a noun and verb form), causality and intention [170–172]. Thompson's framework includes the below risk definitions, however Althaus suggests that to be complete the framework should also include "actual" and "calculated" risk [7].

- *Subjective Risk* – the mental state of an individual who experiences uncertainty or doubt and worries about the consequences.
- *Objective Risk* – the difference between actual loss and expected loss.
- *Real Risk* – the combined probabilities and negative consequences that exist in the real world.
- *Observed Risk* – measurement obtained from constructing a model of the real world.

3 For the thrill-seeker relishing the positive risk felt when BASE jumping from a tall building [64].

- *Perceived Risk* – rough, and potentially under or over exaggerated, estimate of real risk made by an untrained member of the general public.

Such a categorisation highlights how overloaded the term *risk* is. When reflecting real-world event probabilities, *objective* or *real* risk would inherit a quantitative probabilistic quantity. Whereas a *subjective*, *observed* or *perceived* risk projection, made about the outcome or frequency of future events based on prior knowledge, would generally be qualitative in nature.

During a HARA *thought experiment* it is not unreasonable to expect automotive safety practitioners to mix the risk categories described above. Regional accident statistics and injury categorisations are available that quantify *real* risk (e.g. German In-depth Accident Statistics (GIDAS) [175], the Institute for Traffic Accident Research and Data Analysis (ITARDA) in Japan [82], and the Abbreviated Injury Scale (AIS) [98]). These are typically used during the risk classification process (see Section 5.3.3.2 on page 88 and [58]) to estimate the likely severity of an accident scenario. *Subjective* and *observed* risk categories might be in evidence too; particularly when reasoning about the impact that a vehicle system fault might have on the driver's ability to control the hazard (see the discussion on controllability in Chapter 6). Although using real-world models to inform automotive hazard analyses is atypical, simulations and conceptual models (e.g. single and double-track bicycle models [87], MISRA control systems view of a vehicle model (Figure 1 page 19) are frequently used to approximate vehicle behaviour and driver controllability. Superficially, *perceived* risk appears to have little relevance to the automotive safety practitioner's *thought experiment*. That said, having the ability to imagine how drivers might underestimate the risk associated with vehicle technology misuse, might prove insightful for the automotive safety analyst.

Certainly the breadth of understanding, and the implicit ethical commitments [172], that can be read into risk statements, will do nothing to reduce the challenges that exist in communicating AD risks (and benefits) to regulators and to society at large.

4.2 RISK AND UNCERTAINTY

Crucial to today's AD features is the ability to detect objects and events in the environment. Heterogeneous sensor suites and machine learning algorithms are typically used to implement the perception systems required. As discussed in section Section 2.3.4.3, these perception systems are complex and opaque in nature. For the automotive safety analyst this added complexity (see Table 2 on page 49) and potentially introduces uncertainty, but what exactly is *uncertainty* and how does it relate to *risk*? To address this question, this section explores the notion of uncertainty and its relation to risk.

The term *'uncertainty'* is no less unambiguous in its definition than *risk*! Oxford Languages on-line dictionary defining *uncertain* and *uncertainty* as:

- *uncertain* (**adjective**):
  - *" not able to be relied on; not known or definite"* or
  - *"(of a person) not completely confident or sure of something"* [126]
- *uncertainty* (**noun**):
  - *"the state of being uncertain"* [127]

In a literature survey of the media's treatment of *risk* and *uncertainty* Ashe discusses the difficulties associated with defining *risk* and *uncertainty*. In the same way that *risk* can convey the quantitative probability that an event will occur and the subjective qualitative assessment of the magnitude of the harm, the intimation is that *uncertainty* has multiple meanings too.

Ashe suggests that for the general public *uncertainty* could simply mean that "we have no knowledge" about an event. In contrast, for the majority of scientific and engineering disciplines *uncertainty* refers to one of many outcomes that can reasonably be expected in a particular situation. For example, the error or Standard Deviation that can be reasonably expected in a set of measurement data [10].

In contrast, Zinn suggests that *uncertainty* relates to the management of *risk*. In "Social Theories of Risk and Uncertainty" he suggests that *risk* exists in

two forms: Firstly, there is *risk* to do with the management, prevention and decision making associated with *real* or *perceived* current danger or harm, or claims about future dangerous events. Whereas, *risk* in its second form is about the management of *uncertainty* [206]. Aven and Renn agree that *uncertainty* relates to *risk*. Although, rather than advocating that *uncertainty* is a property to be managed, Aven and Renn suggests that risk includes uncertainties: that is, *uncertainty* about the probability of the event occurring and its consequences, and *uncertainty* about the severity of the event and its consequences. As such, the notion of *uncertainty* is not real, but is a "construct of human imagination used to cope with potential future outcomes that can become real" [15].

In a paper focusing on decision making in the presence of risk and uncertainty, De Groot and Thurik cite Knight's 1920's seminal research into economic risk and uncertainty [37]. De Groot and Thurik suggest that unlike *risk*, *uncertainty* includes an element of *chance*. With the distinction being that, although the *risk* of an event might be unknown, the probability distribution is understood. This contrasts with *uncertainty* where both the outcome and the probability distribution are unknown. De Groot and Thurik illustrate the difference between *risk* and *uncertainty* using a dice rolling example. Imagine you are given a die to roll. If the outcome is four or greater you win £50. For an outcome of less than 4 you lose! Assuming the die is unbiased, then the chance of winning or losing is 50%. However, if unbeknown to you the die has been weighted then there is now *uncertainty* regarding the outcome. With the biased die the probability of winning or loosing still exists, however the probability distribution is now unknown.

In their survey of uncertainty, Dutt and Kurian suggests that four classes of *uncertainty* exist, namely:

- *epistemic:* the uncertainty that exists due to a lack of knowledge
- *linguistic:* the lack of precision in human natural language interactions that leads to uncertainty, vagueness, or ambiguity
- *ambiguity:* situations where the probability of the outcome is unknown or where multiple meanings could be understood

- *variability:* situations where there is uncertainty due to internal or external process variation or 'non-static' statistical distributions

When considering the implications for AD feature development and use, in an automotive hazard analysis and risk assessment context, Dutt and Kurian's *uncertainty* classes above provide a potentially useful *thought experiment* framework to explore.

To start the *thought experiment*, *epistemic uncertainty* is undoubtably present. Introducing automation changes the control interaction that exists between the driver and the vehicle, away from a functional safety paradigm (see Section 1.2.2) where driver and vehicle behaviours are consistent and implicitly understood. This segues nicely into the second uncertainty class – *linguistic uncertainty*. As was described at the beginning of Section 2.1, driving is a task familiar to most, which has implications for future automotive hazard analyses. During their work the analyst will typically make assumptions about the driver, the vehicle, and vehicle control interactions. It is unlikely that these assumptions will be explicitly stated, but with the driving experience mutually understood within the team, any implicit assumptions made will be mutually understood and agreed. In an AD context, it would be unwise to rely upon such implicit driving knowledge. As a consequence, greater automation arguably increases the need for increased analysis and specification precision.

Finally, one can imagine that AD vehicle features introduce uncertainty due to *ambiguity* and *variability*. The potential for unsafe vehicle behaviour, due to environmental factors causing AD perception system variability, was a motivation behind the International Standard Safety Of The Intended Functionality (SOTIF) [163]. The same is probably true of *ambiguity uncertainty* also. That is, complex systems operating in complex ODDs leading to potentially uncertain or emergent system behaviours.

One thing that is clear from the above is that the definition of *uncertainty* is not clear! *Uncertainty* does appear to relate to the notion of *risk*. In some contexts, it may refer to *risk management*. However, more typically it conveys the variability that naturally exists when conceptualising the consequences and outcomes of harmful events. That is, the probability of the event occurring, its consequence,

or the severity of the event. Although as De Groot and Thurik suggests, people are typically less sensitive to *uncertainty* than they are to *risk*.

4.3 RISK PERCEPTION AND TECHNOLOGICAL CHANGE

As discussed at the beginning of the chapter, the introduction of AD technologies will increase vehicle capabilities, and with it user expectations. While probably not at the forefront of our minds when we jump in the car, travelling in a passenger vehicle does incur risk. Intuitively, the risk incurred will change as the capability of vehicle systems increases, and the nature of the interactions with the driver and the environment changes. However, in the context of the interaction between humans and automation what affects our perception of risk, and how might our perception change as technology capability and complexity increases?

This section examines the risk perception research literature to inform how the introduction of new AD technologies might influence our perception of risk, and perhaps go some way towards addressing the question "how safe is safe enough?"

4.3.1 *Risk perception*

When choosing an appropriate technological solution it is normal to make a performance versus cost comparison. An arguably equally important social benefit versus cost comparison is far less likely to be undertaken. However, challenges exist that make achieving the goal of maximising the benefits while minimising the costs difficult [150]. For example, once a particular technology becomes a part of the framework of daily life its use becomes very difficult to change. Small pilot studies might provide the means to assess social costs and benefits, but it is atypical for these to be undertaken before

introducing a new technology. Also, consumer technology uptake[4] might outstrip the development time for that technology; meaning that the use of a new technology is widespread before the societal impact of that technology can be fully accessed [150].

Starr's seminal research into technical risk and risk perception used data from historical accidents and health care to make quantitative societal cost evaluations. Comparing these cost evaluations with the assumed societal benefits of a given technology helped build the understanding of "how safe is safe enough?" [150]. Writing in 1969, Starr suggested such "predictive technological assessments are a pressing societal need", but conceded that this would require the development of an objective value system capable of determining quality-of-life improvements. The introduction of the car and the aeroplane provide examples where this quality-of-life cost benefit assessment happened empirically, without any prerequisite "how safe is safe enough?" analysis [150]. The observation that any new technological advancement potentially has unforeseen consequences further supports the need for a more predictive risk assessment framework [55]. However, as discussed above, the potential for both epistemic uncertainty and ambiguity to exist may render the output from such a predictive assessment useless.

The safety engineering practitioner's approach to the risk assessment and safety analysis of new technology is discussed in greater detail in Chapter 5. In contrast to the safety engineer who will likely use a risk assessment technique to consider the impact of new technology, Slovic suggests that the public use an intuitive risk judgement or *risk perception* [140]. One might assume that individuals always have a logical approach to risk. However, Ashe suggests that humans are not "rational calculation machines" where risk is concerned. Instead, Ashe suggests that humans comprehend risk as a mixture of "fact and feeling" [10]. The Royal Academy of Engineering agrees that it would be wrong to assume that an individual always approaches risk rationally, citing an individual's approach to vaccination as an example. The Royal Academy of Engineering suggests that the rational choice, that gives the greatest societal

4 It is interesting to note that Starr identifies this concern at the end of the 1960's, long before the impact of technologies such as the Internet or Smartphones could have been conceived.

benefit at minimum risk to the individual, would be to immunise all children with the MMR vaccine. However, the irrational decision would be for the individual not to vaccinate their child, because the child is being protected by the majority having already been vaccinated [166].

Related to the topic of vaccination, Ashe suggests that *risk perception* gaps also exist – that is, the difference between how afraid one might be of something and the risk that actually warrants [10]. Given that the public's intuitive risk judgement is influenced to a large extent by the media Slovic suggests that an understanding of the public's acceptance of hazards is needed to aid the risk analysis process and to inform policy makers [140]. In a discussion about the challenges of risk and uncertainty media reporting Ashe suggests that to the "lay person" *science* probably comprises a "body of established knowledge". However, for the scientist, once knowledge is sufficiently well established to be considered fact, then it is probably no longer of interest! Instead Ashe suggests that it is *uncertainty* that scientists find most interesting [10].

Discussions regarding societal *risk* and *uncertainty* are probably further muddled by the public and scientists tending to use the two words differently. For the general public, *risk* typically represents a low probability event, whereas *uncertainty* is interpreted as not knowing. Consequently, the ability to report intrinsic uncertainty is an important part of any risk communications[5] [10].

4.3.2 *Risk acceptance*

Probably Starr's most cited research finding is that individuals are prepared to accept risks associated with voluntary activities that are around 1000 times greater than the risks they are willing to accept for involuntary activities. Starr attributes this to individuals using their own value-system to make intuitive

5 Media communications throughout the COVID-19 pandemic perhaps demonstrates this view, with the media often reporting trends, predictions and *a priori* knowledge as *fact*. While changes in expert advice, made as empirical knowledge grew, were often reported by the media as evidence of the Government and its experts making *U-turns*.

judgements about risks and benefits. For example, an individual may choose to move from the city to the country because of the benefits of lower crime rates and better schools, but at the cost of spending more time travelling by car. Contrasting this with involuntary activities, Starr suggested that the criteria and options relating to activities such as natural disasters and electrical power generation, are governed by the applicable controlling body. As a consequence the public perception may be that an amount of rigour has been applied, with a rational analysis of societal benefit versus societal risk having been undertaken - which actually may not be the case [150].

Starr also defined the relationship between acceptable risk and benefit, defining the level of risk that society would deem acceptable as:

$$R_{societal} = V^3_{societal} \tag{1}$$

where $R_{societal}$ is the societal risk and $V_{societal}$ the societal benefit.

Starr suggests that for the individual the "statistical risk of death from disease appears to be a psychological yardstick for establishing the level of acceptability of other risks" [150].

Subsequent research has questioned the validity of characterising risk acceptance on the basis of simply voluntary or involuntary risk. In his paper "The perception of technological risks: A literature review" [35] Covello highlights work that brings into question Starr's original hypothesis. Instead Covello suggests that the public's willingness to accept seemingly greater voluntary risk may be due to other factors. Covello suggest that a *Psychological Paradigm* can be used to understand the relationship between perceived and real risk. This two dimensional taxonomy uses risk factors pertaining to dread and the unknown to determine what the perceived risk will be [140]. In reality activities that are categorised as voluntary may also be considered, by the general public, to be controllable, familiar and non-catastrophic [35].

The Royal Academy of Engineering makes an interesting observation about public attitude to risk, and the irrationalities that can result. They observe

that risks associated with *acts of god* and similar natural disasters are more readily accepted than man-made risks [166]. It is generally accepted that individuals are more acceptant of risk when they are in control or when they have played their part in the decision process, however, The Royal Academy of Engineering observes that history shows that many engineering projects have succeeded because the engineers involved did not include the wider society in their decision making [166].

As touched upon above, the approach to risk used by experts and non-experts is different [10, 35]. Whereas experts tend to use quantitative methods, simulation and experimentation to inform the assessment of risk, non-experts may rely more on qualitative methods, intuition and feelings. The way in which experts and non-experts interpret the available information may also differ. Covello suggests that for the expert no distinction is made between a case where deaths occur together or one at a time over a long period. In contrast non-experts will likely give a higher rating to a single catastrophe that kills many. Typically, non-experts will also apply a higher rating to known deaths (rather than statistically probable deaths) and where suffering is known to occur. For example, the substantial resources that are brought to bear when people have been lost at sea [35].

Medical and rail are two domains where risk reduction and cost benefit comparisons are common place. Although not always explicitly stated, in medicine the trade off between a given treatment's cost, given the finite resources available to healthcare bodies like the National Health Service (NHS), and the potential risk reduction / patient benefit is well known. In their review of the aviation, defence, nuclear, petrochemical and transportation (rail and road) industries, Sujan et al. discuss the notion of As Low As Reasonably Practicable (ALARP) and suggest that the medical domain is unique in its consideration of affordability [155]. Unlike the aforementioned regulated-industries, the medical domain has no common definition of risk. That said, in a medical context the risks and benefits associated with a given procedure are undoubtably very individual judgements anyway.

Individuals are typically more acceptant of risk when the activity is familiar compared to when it is unfamiliar. The activity of driving is certainly one that is familiar, gives the perception of being in control, with often only the individual being exposed to the hazard consequences. Joshi et al. suggests that while drivers still perceive that they are in control and possess the capability to cope then their risk perception will remain low [88]. However, a consequence of this relationship between perceived risk and control might be to change an individual's behaviour. Research suggests that if safety measures move the level of risk below the individual's *risk-thermostat* then they will change their behaviour to reassert a risk level that they were originally content with [88]. For example, might an otherwise cautious driver increase their speed in fog knowing that they have a vision enhancing ADAS as an aid?

4.4 AD IN THE CONTEXT OF RISK AND UNCERTAINTY

In the paper "Risk Objectivism and Risk Subjectivism: When Are Risks Real" Thompson reflects on the objective and subjective nature of risk. Thompson suggests that when considering objective risk it is difficult to be right, and when considering subjective risk it is difficult to be wrong [171]! The literature highlights the lack of a clear definition of risk, with no consistent approach to risk evaluation and risk perception across industries. However, what is clear is that as AD vehicle systems are developed and introduced onto public roads they will be judged by the general public against the backdrop of societal risk.

Much research effort has gone into understanding how people perceive and evaluate risk. While there is general agreement that the public perceive risk and make judgements about risk intuitively [10, 35], there is some disagreement about what influences the way risk is perceived. However, the general agreement is that people are more acceptant of the risk associated with a given activity while the individual feels in control, the activity is familiar, and any consequences appear non-catastrophic [35, 166]. This would explain why as car drivers we readily accept relatively high levels of risk. However, by the same token we will be far less tolerant of the risks associated with AD, where the

perception will be that control has been relinquished to a machine that is more capable than them. Also, any failure that occurs could potentially affect a large number of cooperating vehicles, leading to catastrophic consequences.

Aven suggests that risk has both subjective and objective aspects [13]. Therefore, any risk assessment should include the *frequency probability* of event occurrence together with subjective judgements about likely outcomes based on prior knowledge, and the analyst's assumptions and uncertainties. In contrast, industrial standards have evolved from a more traditional risk model, where accidents are caused by hazards, hazards are caused by a chain of events, which are mitigated by breaking the chain. A road network containing highly complex and collaborating vehicles clearly becomes a system of systems issue, with the potential for emergent behaviours to result from such systems interacting. Normal Accident Theory (NAT), developed by Charles Perrow after the Three Mile Island incident, suggests that accidents are inevitable in complex and tightly coupled systems [63]. Given such an accident inevitability, then undoubtably the risk models used by standards like ISO 26262 should incorporate aspects of the complex non-linear accident model proposed by Hollnagel [79].

Although not a part of today's automotive risk assessment process, undertaking a Cost Benefit Analysis (CBA) as exemplified by governmental regulatory bodies [116, 182], might be prudent in the context of tomorrow's complex AD vehicle features. Sujan et al.'s case study findings suggest that for high severity low probability risk scenarios the application of CBA is problematic; both because a high severity might affect the validity of CBA calculations, plus society might perceive CBA scores differently to the domain experts [155]. However, as Sujan et al. concludes, the risk reduction decision should not be confined to purely ALARP principles. Corporate responsibility, ethical reasoning, potential business benefits and impacts, and simply whether *"it was a good thing to do"* are all important considerations that should be evaluated when considering the introduction of new technology [155].

In conclusion, the risk assessment of a AD vehicle feature becomes a multi-dimensional problem. There is the *real* probabilistic risk associated with the

use of the feature in its environment, but this is not the same as the risk that results from undertaking the analysis. Nor is it the same as the risk perceived by users of the highly automated vehicle feature. Consequently all stakeholders could be using the term risk and meaning very different things.

5

MODELLING, ANALYSING AND MANAGING RISK

Having explored risk's many facets in Chapter 4, this chapter focuses on the engineering aspects of that broader discussion. In doing so, it first describes how engineers, and more specifically safety practitioners, seek to identify, understand and quantify risk. The chapter then considers the approaches to managing risk once identified. Many hazard analysis techniques in use today are decades old. Consequently, they were developed in an era of less complex systems – when candidate systems could be described and reasoned about as a simple black box model. Today's systems are typically complex systems of systems, necessitating the development of techniques like STPA and FRAM. These techniques are discussed in Section 5.2.

This chapter then discusses the notion of joint cognition and joint cognitive systems. This is of significance to this research, because it seeks to address human interaction with automation; in particular human understanding of current events and determination of future actions. The chapter then goes onto discuss some of the challenges associated with analysing complex systems, before finally considering hazard analysis in the context of AD vehicle features.

5.1 ACCIDENT MODELS AND RISK

While discussing the evolution of accident causation models Toft et al. suggest that such models can be categorised as simple sequential linear accident models, complex linear models, or complex non-linear accident models [174].

Simple sequential linear models first emerged in the 1920's. These models assume that a linear sequence of events are required to lead to the initiating

event becoming an incident, with each intervening event acting sequentially on its neighbour to cause the chain reaction. Consequently, the removal of any one event effectively breaks the sequence and thus stops the accident occurring.

Emerging during the 1960's, epidemiological or complex linear models are similar to sequential models. Again, they assume that an accident will be as a result of a combination of events taking place, some of which may be away from the main chain. This has the potential to make some events latent. Acknowledging that events may be latent or remote, allows a broader problem space to be modelled. For example, modelling accident contributing factors outside of the immediate system boundary (such as organisational deficiencies or remote actions).

More recently, complex non-linear accident models have emerged. These accident models have their basis in system engineering principles, suggesting that accidents may be the product of emergent properties of the system. For example, Hollnagel suggests that accidents result from the alignment of conditions and occurrences within the system. Consequently accidents may be caused by everyday system adjustments and not necessarily by a fault or by human error. Toft et al. suggests that it is only by exploring and understanding these subtle interactions between multiple real-world factors that accidents can be fully understood and prevented [174].

5.1.1 *Linear accident models*

SEQUENTIAL MODELS The first accident models were developed to protect workers in industry, and unfortunately assumed that when an accident occurred people were always the culprit. Heinrich's 1931 Domino Model is considered to be the earliest example [80, 94]. The notion here is that any injury can be traced back to its ancestry or social environment which caused the first domino to fall. Once that first domino has fallen, it knocks down its neighbour, which knocks down its neighbour, and so on until the injury occurs.

Thus, by removing any one of the dominos the last injury-causing domino will be prevented from toppling over. So by following this line of thinking, the easiest and most effective approach would be to remove the *unsafe act or condition* domino [94]. Researchers including Leveson are critical of sequential models, highlighting the fundamental flaw that accident symptoms could be identified and removed while the underlying cause remains [94].

The single event Domino Model can be expanded into a chain of events model. Here the accident is broken down into multiple causal factors, each of which has a chain of events leading up to it. Techniques such as Fault Tree Analysis (FTA), event trees and critical path models have historically been used to explore accidents in a systematic way, however Leveson observes that there is often "no real stopping point when tracing events back from the accident" making it difficult to know to what depth the analysis should go [94, 174].

Given that ultimately the initiating event could be the failure of a component within the system, functional safety standards and guidelines typically incorporate a chain of events model within their accident definitions and risk models. As an example, Figure 6 shows the automotive failure model[1] commonly used [113].

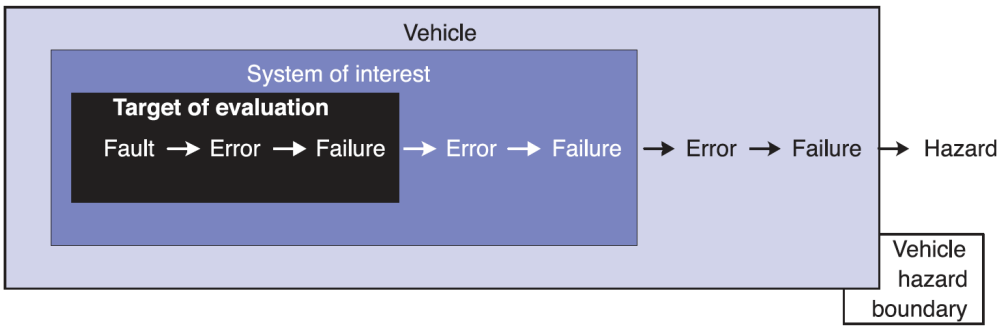


Figure 6: MISRA's fault-error-failure vehicle chain of events (reproduced from [113])

EPIDEMIOLOGICAL MODELS In the 1940's researchers begun describing the complex nature of accidents. These epidemiological models stem from John Gordon's work researching epidemics. He proposed that the agent, the

1 It should be noted that MISRA uses the same definitions of Fault, Error and Failure as IEC 61508 [83] here.

environment and the host are all factors that need to be considered when analysing accidents [94, 174].

This analogy with the effect of pathogens on the human body has been used to describe latent-faults within a system and to describe the human factors considerations associated with latent-errors that may be present when humans use systems. As a consequence accident prevention following an epidemiological model tends to focus on preventive actions and processes to mitigate accidents [174].

SYSTEMATIC MODELS At the end of the 20th Century researchers began to realise that observed accidents were not fitting existing models. Rather than considering accident events in isolation they identified the need to consider accidents in the context of the system's environment in which the accident occurred. Key to understanding this wider context was an understanding of how human error could contribute to an accident [174]. Whereas the traditional accident models laid accident blame firmly with the operator, researchers (such as Reason) acknowledged that latent human errors high up in an organisation could contribute to a given accident occurring. Reason famously represented such an accident sequence as a Swiss Cheese Model – a model containing a number of layers of holey cheese. Without being specific about what each layer of cheese represents, the Swiss Cheese Model shows how even with multiple layers of protection / mitigation within a system an accident can occur. With the holes in the cheese slices representing inadequacies in a system, that if not addressed will line up, and allow inherent hazards in the system to slip through [174].

5.1.2 *Complex non-linear accident models*

Around the same time (1980's) researchers began to talk about complex non-liner accident models. While reflecting on Charles Perrow's NAT model in "Models of causation: safety" Toft et al. conclude that in tightly coupled and

complex systems, if a component fails it is almost impossible to reason about all possible consequences of that failure on the system [174].

Two notable techniques for analysing complex non-linear systems are Systems Theoretic Accident Model and Processes (STAMP) and Functional Resonance Analysis Method (FRAM). Both techniques are discussed in detail in Section 5.2.

5.1.3 *Assessing risk*

In their discussion "Is risk analysis scientific?" Hansson and Aven introduce a 5 step risk analysis model, which they suggest is the process by which domain experts make decisions about risk [70]. Those 5 steps are: evidence, knowledge base, broad risk evaluation, decision maker's review, and decision. Hansson and Aven suggest that domain experts will first gather evidence and build a knowledge base for the given activity. They will then evaluate the knowledge base, giving consideration to the risks and uncertainties involved. Of particular importance are those uncertainties that come into play when one tries to project forward into the future – something that, in an automotive context, the introduction of greater automation will intuitively exacerbate. The final steps of the risk analysis model sees the decision maker use the output from the evaluation, together with any policy considerations, to inform their decision [14]. Corporate reputation, risk responsibility [155] and security [164] are all examples of factors that could influence policy.

Although risk evaluation is typically thought of as a probabilistic and purely quantitive problem, as discussed in Chapter 4, risk has both a subjective and an objective dimension. As previously identified, risk can also be a useful tool to describe a lack of knowledge, as well as knowledge and facts related to the problem itself. For example, when reflecting on past experience, one might consider the risk probability of a hazardous event occurring. In contrast, a hazard analysis activity will include uncertainty with the analyst using engineering judgement to reason about the likelihood of the event occurring

with little or no prior experience. To illustrate this let us consider the Chernobyl accident [196]. Prior to the accident itself one might expect the risk assessments for such a disaster to be purely subjective; including conjectures about accident outcomes. However, as a consequence of the accident on 26th April 1986 *real* risk data begins to emerge making an objective assessment tenable.

In an automotive context objective and subjective aspects of risk can be found within the ISO 26262 risk model [162], although the explicit distinction is not made. Within ISO 26262 hazard risk is a function of the severity of the accident, the probability of being exposed to the hazardous event and its controllability (see Chapter 6 for an in depth controllability discussion).

$$R = f(F, C, S) \tag{2}$$

where F = the frequency of occurrence, C = the controllability, and S = the severity of the resulting harm.

In cases where accident data or reliability data is available then an objective measure for severity and probability of exposure may be possible. More typically, insufficient data is available for such an objective assessment, so engineering judgement will be relied upon when classifying severity and probability of exposure.

Obtaining objective controllability data is also a non-trivial task [33, 45]. As a consequence, like severity and probability of exposure, automotive risk assessment will typically include a subjective controllability assessment. For manually controlled vehicles the uncertainty associated with such subjective assessments are likely to be small; with significant engineering knowledge having been accrued over decades. However, the introduction of greater vehicle automation changes the relationship between the driver and the vehicle, and hence the uncertainty associated with subjective controllability assessments. Therefore, following Aven and Renn's line of reasoning, to adequately express automotive risk a future automotive risk model should address uncertainty explicitly.

5.2 ANALYSING COMPLEX AND JOINT-COGNITIVE SYSTEMS

5.2.1 *Complex system analysis*

Leveson suggests that the event chain model (see Section 5.1.1) constrains thinking in a number of ways: Firstly, there is a tendency to think about the accident sequentially – with the accident only resulting from some initiating event like a component failure, and with there being no opportunity for feedback to the cause. Secondly, being sequential can constrain the way one thinks about countermeasures. Finally, accidents that occur as a result of nothing going wrong would simply be out-of-scope of causal chain models [95].

STAMP is a model or framework that applies Systems Engineering thinking to accident analysis by acknowledging that safety is an emergent property of a system. That is, the notion that the behaviour of a complex system (comprising multiple subsystems) will not behave simply as the sum of its parts. Instead, a complex system will exhibit emergent properties that make analysing individual components in isolation ineffective at identifying all error states. With emergent properties arising from the behaviour of, and the interactions between subsystems, Leveson and Thomas argues that any analysis technique should also consider these interactions. Viewed in this way safety becomes a dynamic system property that has to be understood and managed if accidents are to be avoided [96].

Two STAMP based tools exist: Systems-Theoretic Process Analysis (STPA) is an analysis technique that analyses the potential causes of accidents during development, thus allowing hazards to be eliminated and controlled during development. While Causal Analysis based on Systems Theory (CAST) examines accidents and incidents that have occurred to identify the causal factors that were involved.

5.2.1.1 *STPA*

As described above, STPA is a safety analysis technique based on STAMP. STPA as described in the *STPA Handbook* comprises the four steps as shown in Figure 7 [96].

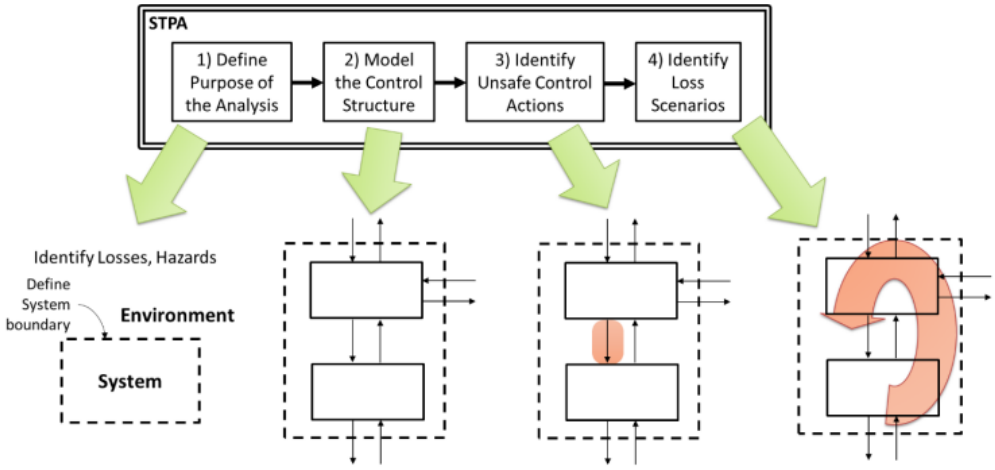


Figure 7: The STPA process (reproduced from the *STPA Handbook* [96])

STPA Step 1 defines the purpose of the analysis. That is, what are the losses that the analysis seeks to prevent? The *STPA Handbook* defines a *loss* as "something of value to a stakeholder" [96], with the term *loss* being chosen to make the notion domain agnostic. In an automotive context, a *loss* is best interpreted as an accident. From the losses, hazards are then derived. STPA defines a *hazard* as "a system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to a loss" [96]. The last activity within Step 1 is to define system-level safety constraints for the system under consideration. These are effectively the highest-level safety requirements for the system that describe the conditions or behaviours that the system shall exhibit to prevent hazards.

Having defined the scope of the analysis, Step 2 of the method then builds a model of the system – referred to as a Control Structure Diagram (CSD). STPA defines an CSD as a hierarchical control structure comprising feedback control loops. By modelling a system's interrelationships and interactions as control loops between subsystem elements, the control structure needed to constrain the overall system behaviour is described. This system model forms the basis

of all subsequent analyses. Stanton suggests that with its hierarchical structure the CSD is suited to modelling technical and perhaps environmental failures. However, he suggests that the CSD is less suited to modelling complex human or organisational interactions because "the relationship between agencies and agents are not necessarily hierarchical" [147].

As highlighted by Leveson and Thomas, a challenge faced by analysts undertaking hazard analysis is managing complexity [96]. They suggest that abstraction should be used to manage such complexity. As such, an initial CSD should start at a high level of abstraction, and then each future iteration should add detail. However, given the complexity of some automotive Control Structure Diagrams[2] seen in the literature, this challenge is real and often not addressed well in practice. Additionally, the knowledge and time needed to construct the CSD should not be underestimated [148]. As will be discussed in Chapter 9, using the EVCM as the basis for the CSD, contributes to supporting the use of STPA in an automotive context.

Systems will typically comprise multiple control loops and controllers. Being a hierarchical control structure, the controller with the highest authority will be placed at the top of the CSD, with the controllers below having progressively less authority over the system. This means that typically the human in the system (e.g. the pilot [31], or the driver [2, 4]) is designated as the top controller within the hierarchy. This research postulates that by modelling the human as such, there is a tendency for interactions between the human and the automation to be viewed as traded control, where the human and the machine take turns being active (see Section 2.3.3). This might be appropriate in situations where the human operator is simply monitoring and responding to automation state changes, but it does seem somewhat artificial when shared continuous control is involved – as in the case of driving.

The third STPA process step then analyses how the control actions in the CSD could lead to the losses identified in Step 1. Those control actions identified as having the potential to cause losses are referred to as Unsafe Control Actions

2 For example, the Control Structure Diagram presented in *Functional Safety Assessment of an Automated Lane Centring System* contains 26 individual system elements [21].

(UCAs). Similar to techniques, such as Hazard and Operability Study (HAZOP), STPA uses guide words to aid the discovery of UCAs in a systematic and structured way. Various guide words are used, but they seek to address the following broad categories:

- How might *not providing* a control action in a given context lead to a hazard?
- How might *providing* a control action in a given context lead to a hazard?
- How might providing a control action with the *wrong timing* in a given context lead to a hazard?
- How might providing a control action with the *wrong duration* in a given context lead to a hazard?

Step 3 also involves defining *controller constraints*, where "a controller constraint specifies the controller behaviour that needs to be satisfied to prevent UCAs" [96]. In this regard, controller constraints can be thought of as safety requirements supporting the system-level safety constraints defined in Step 1.

The final step (Step 4) then involves identifying how the identified UCAs might occur. The *STPA Handbook* describes two scenarios in which UCAs will occur: The first scenario being where the controller itself has gone wrong, something has gone wrong with the feedback to the controller, or the controller has received incorrect feedback from another controller. The second scenario is where the controlled process, or plant, has not responded correctly to a control action. The *STPA Handbook* includes questions and illustrations to help the analyst visualise the two scenario types and thus uncover pertinent loss scenarios. That is, the scenarios (or sequence of events) in which a particular control action would become unsafe.

When the potential exists for a control action's outcome to be influenced by multiple system actors (e.g. the human and / or the automation) then consideration should be given to cognitive principles too. Enhancements that have been made to STPA in relation to shared control are addressed in Section 5.2.2.

5.2.1.2 *FRAM*

As discussed above, the typical approach is to develop a model that describes the accident sequence, or structure of the system (effectively a hierarchy of its layers, parts and components), and then to apply an analysis method to this model. This approach is referred to as a *model-cum-method* approach [80].

Hollnagel suggests that a *method-sine-model* approach is better than a *model-cum-method* approach. This is because a method is used to explore the functionality of the system, from which the system model is then derived. It is suggested that the advantage of this approach is that it avoids constrained thinking; as no system model or assumptions are made before the analysis begins, unintended constraints in thinking become less likely [80].

FRAM assumes that a system will be subject to system variances and tolerances. Examples of which could include variances in the environment in which the system operates or variances in human behaviour. Once these variances and tolerances become too large then the system is no longer able to absorb them and an accident results. Within the FRAM methodology this is referred to as *functional resonance*.

FRAM is cited in relation to this research because it was considered as a candidate analysis technique for the EVCM. However, with its use as an analysis technique being atypical in the automotive industry, the decision was made to not pursue it as a technique for this research. As a future research topic, using FRAM to explore an AD system's robustness to shared control variances may prove fruitful.

5.2.2 *The notion of joint cognitive systems (JCS)*

In the context of shared control, intuitively time is important. With many contemporary research activities in the area of AD discussing timing implications in the context of: a driver's resumption of control [50], driver take-over quality [40], driver decision making [51], and impact of system failures on driver

interactions [69]. Consequently, a discussion on social-technical systems and specifically Cognitive Systems Engineering (CSE) research and the notion of Joint Cognitive Systems (JCS) is pertinent.

A cognitive system is defined as "a system that can modify its behaviour on the basis of experience so as to achieve specific anti-entropic ends" [81, page 22]. As such, the activity of driving can be viewed as a JCS; as a combination of actions are required, within the context of a dynamic and sometimes unpredictable environment [81]. A literature search identified CSE as an area of research *in vogue* in the 1980's, with researchers like Hollnagel and Wood publishing a number of papers and books during that decade. However, with the exception of the medical domain, little has been said about JCS since the 1980's. Hollnagel himself [76] suggests that the demise of CSE is due in part to how the phrase is parsed. Hollnagel's intention was for the focus to be *[cognitive systems]* engineering, or the design and development of joint cognitive systems. Instead Hollnagel suggests many researchers consider cognitive *[systems engineering]*, or the knowledge underpinning systems engineering. Perhaps somewhat disgruntled by this apparent misinterpretation, Hollnagel appears to have turned his attention to system resilience and techniques such as FRAM. This is unfortunate, because CSE and the notion of JCS do include a framework that shows promise for reasoning about the temporal aspects of HMI.

5.2.2.1 *COCOM*

Rather than considering information processing performance in relation to joint cognitive systems, Hollnagel and Wood suggest that the "regularities of performance" should be considered instead. The Contextual Control Model (COCOM) is a cyclical model of human action, based on three concepts: *competence, control* and *constructs*. With *competencies* being the possible actions that can be taken, *control* being the way in which *competencies* are chosen, while *constructs* describe the context in which the action is carried out [81]. Control is then broken down into the subcategorises of: Scrambled, Opportunistic, Tactical, and Strategic. At one extreme scrambled is at the "hit and hope" level of control, while (as the name suggests) strategic is non-time dependent

high-level goal based control. Hollnagel and Wood suggest that humans tend to operate somewhere between opportunistic and tactical control, using a combination of feedback and feedforward decision making. Hollnagel and Wood observe that humans tend to "shy away" from strategic control because that requires effort! In an AD driving context, this would support the notion that anecdotally drivers tend to develop their understanding of how a vehicle's automation operates through experimentation rather than through reading the Owner's Handbook.

Hollnagel and Wood suggest that effective control requires the human or system to evaluate events correctly and then to choose the appropriate actions. In this regard, Hollnagel and Wood has been openly critical of HMI designers, suggesting that because designers focus on observable behaviour, the concept of time has been largely ignored. Within the COCOM this "cyclical model of human action" incorporates two control loops (within the red arc in Figure 8) to mirror this notion: *event evaluation* and *action selection*. Important for *event evaluation* is the time needed to evaluate events (T_E), while the time needed to select an action (T_S) is important in the *action selection* loop. The time needed to accomplish both the event evaluation (T_E) and the action selection (T_S) must then be viewed in the context of the time available (T_A), as well as the time needed to actually complete the action (T_P) (the yellow arc in Figure 8).

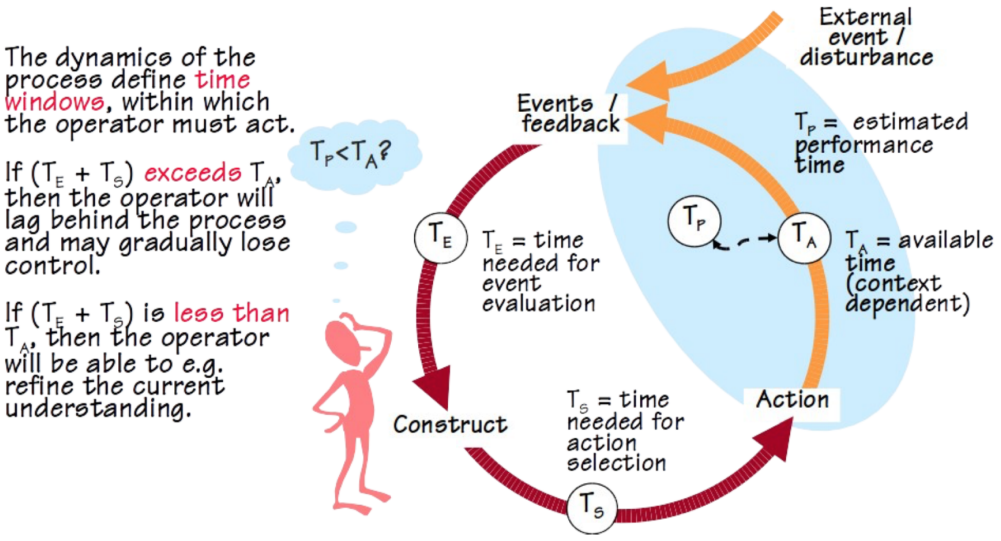


Figure 8: The "cyclical model of human action" (COCOM) (reproduced from [77])

Although contemporary research now suggests that the human cognition actually uses a process of *predictive processing* [48], that is continuously predicting its own sensory inputs and acting upon the predictive error, the *competency, control, constructs* loop at the heart of COCOM is still likely to have utility.

5.2.2.2 *ECOM*

The Extended Control Model (ECOM) extends COCOM by introducing a number of performance control loops – with each control loop within a JCS happening in different time frames. For example, some will be closed-loop and reactive, others will be open-loop and more proactive in nature, while some will be a mix of the two [78]. Although the performance levels are not fixed, the following are suggested and do seem appropriate in an automotive context: *tracking, regulating, monitoring* and *targeting*.

In this context the *tracking* layer deals with automatic corrective actions that are purely feedback in nature, and result from external disturbances. As such, it would be appropriate to consider these to be analogous to the Control Level in Michon's hierarchical model (see Figure 2 page 27). Making steering corrections to maintain the vehicle's heading following a wind gust, or modulating the accelerator pedal to achieve the desired vehicle speed would be driving examples of *tracking* control loops. The *regulating* layer uses predominately feedback control, with some feedforward control. Example *regulating* layer functions could include changing speed relative to other road users or manoeuvring around obstacles. The *monitoring* layer is largely feedforward in nature, and is about setting objectives and activating plans. *Monitoring* tasks might include monitoring the status of the vehicle (e.g. monitoring vehicle speed), and progress towards the required destination. In an ADAS context, the monitoring of the automation would fall into this performance layer. The *regulating* and *monitoring* layers are analogous to Michon's Manoeuvring Layer tasks. At the top of the ECOM model is the *targeting* performance layer. This layer is analogous to Michon's Strategic Level and sets the goals for the given journey. Being concerned with high level goal

setting, the *targeting* layer tends to be purely open loop, although it is informed by feedback from the lower layers. For example, route planning might use land marks identified by the monitoring layer to estimate progress towards a given destination.

Within the ECOM the different performance layers operate simultaneously, while the actions at one layer may or may not affect other performance layers. For example, a pedestrian walking out into the road might affect the *tracking* layer, but would not affect the layers above. In contrast, the loss of localisation (i.e. getting lost when driving in an unfamiliar city) might impact *regulating* and *tracking* also – because as drivers we tend to slow down when we are not sure where we are.

5.2.3 *Engineering for humans*

As mentioned in Section 5.2.1.1, a criticism levelled at STPA by this research[3], is the tendency to elevate the human to the top of the control hierarchy, which does not adequately conceptualise the nature of shared control. In her MSc thesis "Engineering for Humans: A New Extension to STPA" France addresses this to some extent by incorporating cognitive principles (i.e. competency, control, construct) into the "human" block in the CSD [57]. France considers driving examples in her thesis, so the "human" block is the driver. However, reflecting again on the task of driving in the context of Michon's HCM or Hollnagel's ECOM, maintaining the driver at the top of the hierarchy may still fail to uncover hazard causes due to the shared nature of driving tasks – particularly in regard to driving tasks / competencies that are occurring at the control (or tracking), regulating or monitoring levels.

Building on COCOM, France identifies further scenarios where UCAs might lead to losses occurring. Thus STPA Step 4 is embellished by asking the following questions [57]:

- How did the operator choose which control action to perform?

3 Having observed how CSDs are typically constructed.

- What does the operator know or believe about the system?
- How did the operator come to have their current knowledge or beliefs?

Although France's research is aimed squarely at the human agent in a JCS, incorporating the notion of perception, knowledge and control action selection (see discussion on SPA and SUDA in Section 2.3.4) does raise the question as to whether the same principles could be applied to other system agents. For example, could these principles be used to reason about system agents that implement learning algorithms such as neural networks?

5.3 RISK MANAGEMENT

The field of risk management is considered relatively young by some (e.g. Aven in [14]) because research literature on the subject did not appear until the latter half of the Twentieth Century. As a discipline, risk management can be considered as having two distinct endeavours: the first is to understand and manage the risks associated with a particular activity, while the second is a generic risk research and methods development. Viewed against this paradigm, this thesis seeks to contribute to the *methods development* aspect of risk management, but for the specific *activity* of AD when vehicle control is shared.

Like all engineers, automotive engineers are happiest when dealing with discrete definitions and quantifiable parameters that can be measured! But, as Chapter 4 highlighted, risk's definition is not atomic and typically risk assessments are qualitative in nature – thus they include an amount of subjectivity and uncertainty. As a consequence, risk management becomes subjective in nature making the question *"is it safe?"* a difficult question to answer. This is particularly problematic in the automotive sector [152]. Here, multiple OEMs and suppliers are vying for business in a litigious consumer driven market, so having a common risk measure would be advantageous. Aven suggests that to stand on a strong scientific footing, risk management should use a common set of definitions. However, Aven does concede that

due to the subtle differences in risk perception across industries that goal is probably unachievable. The potential consequence of which is to impact how risk is understood, analysed, managed and reasoned about [14]. The introduction of ISO 26262 has certainly helped normalise the automotive risk assessment and management vocabulary, but new vehicle features and technologies will undoubtably challenge society's understanding of what constitutes *safe*.

5.3.1 *Risk management strategies*

The strategy behind risk management is the reduction of risk, for which a combination of three strategies are proposed [14]. The first is a *risk informed strategy*. This strategy relates to the treatment of identified risk, to avoid, reduce, transfer, or control each risk identified. The second strategy is a *cautionary / precautionary strategy*. This approach is built on the principles of redundancy and resilience and uses factors such as containment, safety factors and safety device redundancy. The third strategy is a *discursive strategy*. This approach involves building confidence and trust in a given system, achieved through engaging all stakeholders to reduce uncertainties and ambiguities, clarify facts, allow deliberation and promote accountability.

Aven stresses the difference between cautionary and precautionary risk management strategies, noting that a precautionary risk management strategy should normally prevail, as once identified action should be taken to mitigate that risk. For example, the Norwegian oil and gas industry routinely protecting living quarters with fireproof material as a precaution [14]. With their acknowledgement of uncertainties in system modelling and the need to develop a system's hazard resilience, both STAMP and FRAM can be thought of as examples of precautionary risk management strategies [14].

5.3.2 *When to stop*

Ultimately risk management is a balancing act, where profits need to be balanced against safety as well as factors such as reputation and societal acceptance. Typically constraints will be applied to the balance, and referred to as *risk criteria*, *risk acceptance criteria* or *tolerability criteria* [14].

A *risk acceptance criterion*, at the foundation of UK Health and Safety Law, is the concept of As Low As Reasonably Practicable (ALARP). The definition was set out by the Court of Appeal (in its judgment in Edwards v. National Coal Board, [1949]) [72] as:

> 'Reasonably practicable' is a narrower term than 'physically possible' ... a computation must be made by the owner in which the quantum of risk is placed on one scale and the sacrifice involved in the measures necessary for averting the risk (whether in money, time or trouble) is placed in the other, and that, if it be shown that there is a gross disproportion between them – the risk being insignificant in relation to the sacrifice – the defendants discharge the onus on them.

Or to paraphrase into a single sentence "risk-reducing measure shall be implemented unless it can be demonstrated that the costs are in gross disproportion to the benefits gained" [14].

Given its place in UK Law, industries such as aviation, defence, nuclear, petrochemical, road and rail are all regulated by standards that address the ALARP principle. As a consequence being able to justify that risk has been reduced ALARP is an overarching principle. That said, arguing sufficiency is not purely an ALARP argument. In practice factors such as corporate responsibility, ethical reasoning, potential business benefits and impacts will all influence the stakeholders decision making process [155].

5.3.3 *Relevant safety standards*

Each industry sector will have applicable standards guiding the application of system safety for that sector. Too numerous to mention here, this section focuses on the three international standards pertinent to the automotive sector: IEC 61508 [83], ISO 26262 [162] and ISO 21448 [165].

5.3.3.1 *IEC 61508*

In 1978 the International Electrotechnical Commission (IEC) began development of the engineering standard IEC 61508. The aim of which being to describe a process whereby the functional safety[4] of a system and its components could be assessed and achieved. Herrmann suggests that the committee set up to create IEC 61508 were the first to acknowledge safety as a systems issue. Also, that there was a need to harmonise the safety management process across all system components irrespective of the technologies used. Although written as a generic standard, the majority of the committee came from the nuclear and offshore industries [74].

IEC 61508 comprises seven parts that describe principles, techniques and measures for the functional safety lifecycle phases from the initial concept development through to decommissioning. The fundamental principle is to determine the risk associated with a given Equipment Under Consideration (EUC) and then to deploy safety functions having the required safety integrity level[5] to achieve or maintain safety. Those safety functions will likely be implemented by an E/E/PE system, but could be achieved using some other risk reduction measure. Being generic in nature, it also forms the basis from which other industry specific standards can be derived.

4 IEC 61508 defines *safety* as "This is freedom from unacceptable risk of physical injury or of damage to the health of people, either directly, or indirectly as a result of damage to property or to the environment" and *functional safety* as the "part of the overall safety that depends on a system or equipment operating correctly in response to its inputs [83].

5 IEC 61508 defines *safety integrity* as "the probability of an E/E/PE safety-related system satisfactorily performing the specified safety functions under all the stated conditions within a stated period of time."

To determine the safety integrity all system failures, capable of leading to an unsafe system state, are considered; that is, random hardware failures, software induced failures and failures caused by electrical interference. For hardware failures, whose probability of failure can be quantified, either the frequency of the dangerous failure, or the frequency of the safety function failing to operate on-demand can be determined. Systematic failures due to software or development errors are difficult to quantify, consequently the notion of safety integrity comprises both a quantitive and qualitative aspect [83]. Therefore to comply with the standard the implemented safety function will have both reliability targets and process measures prescribed. The higher the integrity level the more stringent the reliability and design rigour requirements become.

The standard describes five methods for determining the Safety Integrity Level (SIL) requirements: The ALARP method, the quantitative method of SIL determination, the risk graph method, the Layers of Protection Analysis (LOPA) method, and the hazardous event severity method. Irrespective of the methodology chosen, the underlying principle is to reduce the risk in the system to a tolerable level. The automotive industry typically uses the risk graph method (see ISO 26262 discussion below) to reason about risk, contrasting with ALARP being the basal principle in UK Health & Safety Law.

5.3.3.2 *ISO 26262*

ISO 26262 was first published in 2011 and revised in 2018 [162]. It is an adaptation of IEC 61508 [83] written to address the specific needs of applying the functional safety lifecycle to road vehicle Electrical/Electronic (E/E) systems: "This adaptation applies to all activities during the safety lifecycle of safety-related systems comprised of electrical, electronic and software components" [162].

Although based on IEC 61508, the emphasis of ISO 26262 is subtly different. While the IEC 61508 functional safety lifecycle emphasises the deployment of electronic systems to reduce the risk associated with system hazards,

the ISO 26262 functional safety lifecycle emphasises addressing the hazards that may result from the malfunctioning behaviour of the E/E system. The consequence of this is that in an automotive context functional safety becomes a system attribute or property.

The Standard makes explicit statements about its scope – about what it does, and does not include:

> ISO 26262 addresses possible hazards caused by malfunctioning behaviour of E/E safety-related systems, including interaction of these systems. It does not address hazards related to electric shock, fire, smoke, heat, radiation, toxicity, flammability, reactivity, corrosion, release of energy and similar hazards, unless directly caused by malfunctioning behaviour of E/E safety-related systems.

This is significant for the development of AD vehicle features as ISO 26262 also states that:

"ISO 26262 does not address the nominal performance of E/E systems, even if dedicated functional performance standards exist for these systems (e.g. active and passive safety systems, brake systems, Adaptive Cruise Control)" [162].

Like IEC 61508, ISO 26262 also includes the notion of integrity levels. But, perhaps a little confusingly, ISO 26262 uses the term Automotive Safety Integrity Level (ASIL) in two ways, with two subtly different meanings. During the *Concept Phase* ASIL indicates the level of risk a particular hazardous event has, while during the *System Design* phases ASIL is used to indicate the level of design rigour required.

The hazard analysis and risk assessment activity, undertaken during the *Concept Phase*, classifies the ASIL for each hazardous event by considering the probability of being exposed to that hazardous event, the probability of being able to control the hazardous event in order to avoid an accident, and the likely severity of the accident should it occur. The matrix in Figure 9 is then used to combine these three subjective measures into an ASIL rating of QM, A, B, C or D, with ASIL D indicating the highest risk. QM stands for Quality

Management, and represents a low level of risk which can be addressed during the development by normal quality management activities.

Severity class	Probability class	Controllability class		
		C1	C2	C3
S1	E1	QM	QM	QM
	E2	QM	QM	QM
	E3	QM	QM	A
	E4	QM	A	B
S2	E1	QM	QM	QM
	E2	QM	QM	A
	E3	QM	A	B
	E4	A	B	C
S3	E1	QM	QM	A
	E2	QM	A	B
	E3	A	B	C
	E4	B	C	D

Figure 9: Automotive safety integrity level (ASIL) determination (from [162])

Once the hazards associated with the E/E system have been identified, and their risk determined, the next step is to define the safety goals for the system. The safety goals represent the highest level requirements and describe the behaviours of the system that relate "to the prevention or mitigation of the associated hazards in order to avoid unreasonable risk" [162]. Each safety goal inherits its ASIL from the hazards it addresses.

Through the *System*, *Hardware*, and *Software* development phases of the safety lifecycle (ISO 26262 Parts 4, 5 and 6 respectively) the ASIL then takes on its second role - as an indication of the *design rigour* required (aligning with IEC 61508's definition of safety integrity). Within ISO 26262 Parts 4, 5 and 6 the processes and methods needed to reduce the likelihood of systematic failures are described; with the recommendation for use (highly recommended, recommended, or no recommendation for or against its use) being indicated by the ASIL [162].

In an ISO 26262 context functional safety is achieved when hazard risk caused by the malfunctioning behaviour of the E/E system has been reduced to a reasonable level. As discussed above, ASIL QM represents a level of system risk that can be addressed adequately using contemporary quality management

practices. Therefore, it follows that QM can be regarded as representing a risk level that society considers *reasonable* in an automotive context. Thus, if vehicle sign-off testing demonstrates that the Item[6] behaviour can be assessed as QM, with the safety mechanisms implemented, then the residual risk is considered *acceptable*. Thus, functional safety has been achieved and the product can be released for production.

5.3.3.3 *ISO 21448*

As discussed in Section 5.3.3.2, the ISO 26262 scope states that it addresses *malfunctioning behaviour* of E/E systems. Some automotive practitioners interpret *malfunctioning behaviour* as an all-encompassing term that covers any *unintended* system behaviour (e.g. component faults, software / systematic failures, functional insufficiencies, performance limitations). However, many interpret *malfunctioning behaviour* purely as component faults and systematic failures. With the introduction of ADAS and AD systems this difference in interpretation was seen as a problem – particularly for systems whose Situation Awareness (SA) is critical to safety. As a consequence, in 2019, ISO published the Publicly Available Specification (PAS) entitled Safety Of The Intended Functionality (SOTIF). This has since been superseded by the International Standard ISO 21448 [165] which

> is applicable to intended functionalities where proper system situational awareness is essential to safety and where that situational awareness is derived from complex sensors and processing algorithms, especially functionalities of emergency intervention systems and systems having levels of driving automation from 1 to 5.

In doing so, the Standard acknowledges that the safety of today's AD systems cannot be achieved by simply drawing a boundary around a system and analysing what lies within.

6 ISO 26262 defines an Item as "system or combination of systems, to which ISO 26262 is applied, that implements a function or part of a function at the vehicle level" [162].

With the *specification and design* available, the first analysis step in the SOTIF process is to identify any potential sources of harm. That is, to identify the triggering conditions that could lead to hazards and to evaluate the risk of the identified hazards. The process suggests prior knowledge, field data and lessons learned as possible means of identifying triggering conditions [165], while for risk classification the ISO 26262 risk schema is proposed [162]. This approach still requires the automotive safety analyst to conduct *thought experiments* to identify likely hazards and to reason about hazard risk. However, with outcomes being influenced by the vehicle's environment and the shared nature of the driving task, effective hazard risk reasoning becomes complex.

5.4 INFLUENCE OF AD ON MODELLING, ASSESSING AND MANAGING RISK

Being focused on the malfunctioning behaviour of vehicle features, automotive safety engineers have typically used techniques such as Functional Failure Analysis (FFA) and HAZOP studies to identify hazards and hazard causes associated with the vehicle feature under investigation. More recently STAMP, and specifically STPA has gained popularity in the automotive industry.

Rather than decomposing a system into subsystem elements and focusing on internal hazard causes (as typified by FFA and HAZOP) [96], STPA requires the analyst to consider interactions between subsystem elements, human operators, and environmental factors that could lead the system to generate unsafe control actions. This broader view of hazard causation is perhaps why STPA is the popular choice when analysing AD features where the need to consider the safety impact of external factors is important [2, 21, 101].

The addition of the SOTIF standard certainly enhances the automotive standards guidance in relation to AD vehicle feature development. However, acknowledging that AD feature hazard causes are due in part to the system's SA [165] drastically increases the scope of any risk assessment undertaken. This po-

tential combinatorial explosion combined with a broadening scope are seen as real challenges for the automotive system safety practitioner (see Section 3.1.1 regarding effective Hazard Analysis and Risk Assessment). The effective use of abstraction is clearly helpful in this regard (see Section 5.2.1.1) further supporting an argument to use STPA to analysis AD vehicle features.

With the SOTIF standard also acknowledging that causes of hazardous behaviour could also include "the unexpected behaviour due to decision making algorithm and / or divergent human expectation" [165] the need to view an AD vehicle as a joint-cognitive system is evident. Therefore, until we reach a time when highly automated driving prevails (i.e. SAE Level 4 and above, where the safety of the vehicle does not rely on the human having some level of safety oversight) then shared control as a potential hazard cause should be considered.

In the context of risk management, knowing *when to stop* is an important criterion. For the traditional automotive system (following an ISO 26262 development approach) this decision would be based on an argument that all system internal faults, capable of causing hazards, have been mitigated such that the residual risk is acceptable. For an AD vehicle feature, whose safety is influenced by correct SA, then hazards that result from insufficiencies in the system's intended behaviour must also be addressed. SOTIF introduces the notion of acceptance criterion[7]. So, not only does the *when to stop* decision for an AD vehicle feature need to balance cost and timing against the acceptance criterion, but being a commercially available product it should also be defensible from a product liability perspective. Although a hugely important research question, the question of *when to stop* is considered too large to address effectively within this thesis. However, a robust argument that shared control has been fully explored, in the context of AD vehicle feature development, is of relevance to this thesis (see Chapter 10), and does provide an important contribution toward knowing *when to stop*.

7 SOTIF defines the acceptance criterion "representing the absence of an unreasonable level of risk". The acceptance criterion can be qualitative or quantitative – e.g., a specific hazardous behaviour, an hourly incident rate, or ALARP – and is expected to be confirmed during the verification and validation programme [165].

6

CONTROLLABILITY

When developing a new automotive feature, or adding an existing feature onto a new vehicle platform, consideration must be given to any new hazard events (as described in Section 5.3.3.2). When hazard risk is determined using the ISO 26262 standard, the severity of the impending accident, together with the probability of exposure to and the controllability of the hazardous event are assessed to determine the ASIL [162]. With all hazardous events classified, the activities undertaken to develop and integrate the feature can then be tailored to the highest ASIL identified; with the highest integrity level demanding the highest design rigour [65].

The automotive industry relies heavily on subjectivity when considering controllability, as the controllability rating for a hazardous event often relies on domain experts combining their knowledge of the feature behaviour with their experience in vehicle handling. Expressed in this way, the assessment of controllability may seem somewhat *un-scientific* and lacking rigour. However, when considered in context – a mature industry that has slowly evolved its products over decades, and where the user interface has remained relatively stable for a century – a subjective engineering judgement and justification based assessment is viable.

The introduction of AD vehicle features challenges such a gradual product evolution paradigm. The increased automation not only changes the driving task and the way in which the driver interacts with the vehicle, but more complex automation also change the way products are developed; particularly the implicit assumptions made about the driver and their ability to control the vehicle should a failure occur.

In this chapter the origins of controllability are discussed, together with the assumptions made about the driver, their role in the control loop and the driving task itself. Familiar vehicle systems are used to illustrate how the notion of controllability works when the driver is viewed as an intrinsic part of the control loop as is the case for a fully manual vehicle (i.e. SAE Level 0). Section 6.3 then introduces two pivotal human factors papers and explores what the literature tells us about how greater autonomy may affect the notion of controllability. Again emerging AD features are used to illustrate the points raised.

6.1 WHAT IS CONTROLLABILITY?

6.1.1 *Origins of controllability*

Mathematical control theory describes the notion of controllability as the ability to steer a dynamic system from an initial to a final state using admissible inputs [89], with the notion tracing back to the work of Kalman in the 1960's [22].

From a vehicle handling perspective, controllability refers to the relationship between the vehicle's fundamental motion characteristics with the driver's personal assessment of the vehicle's handling quality [3]. In the aerospace domain various studies have been undertaken to facilitate the qualitative evaluation of the pilot's subjective assessment. One notable pilot-rating scheme is the *Cooper Harper Rating Scale* devised in the late 1960's to help pilots and engineers evaluate aircraft handling and stability [34].

The notion of controllability can also be found within the commercial and military aerospace standards. The military standard MIL-STD-882C [181] included the concept of controllability in its software control categories. These could be thought of as a sliding scale of software criticality: with *Category I* software exercising full autonomous control over the system (with no potential for intervention to mitigate the hazard), *Category IIb* software providing display information needing immediate operator action, through to *Category*

IV (software having no safety critical control functionality). Inspection of the commercial aviation standards finds a similar notion of controllability in use. For example DO-178C [41] includes system failure categories that consider "the flight crew's ability to cope", from which software levels are derived.

From an automotive system safety perspective the notion of controllability traces back to the Dedicated Road Infrastructure for Vehicle Safety in Europe (DRIVE) projects undertaken in the early 1990's [85]. Here controllability is the probabilistic attribute linking a hazardous event to an accident; by indicating how likely the person in harm's way will control the hazardous situation and thus avoid harm. So for a hazardous event categorised as *nuisance only* the inference is a zero probability of that event becoming an accident. Whereas, a hazardous event categorised as *uncontrollable* has no positive outcome that can be influenced by human intervention.

The DRIVE team categorised controllability into five levels (Figure 10) enabling its relationship to integrity levels to be considered. Although appearing to link controllability to integrity level directly, it should be noted that the risk model used by publications, such as the MISRA Guidelines [113] and ISO 26262, [162] actually link the probabilistic attributes of the hazardous event (exposure and controllability), with the likely severity of the resultant accident [114].

The hazardous nature of driving complicates the classification of hazards because it is often possible to conceive numerous outcomes from the same hazardous event – at the extreme even the identification of a fatal outcome may be possible from the most benign hazardous event [86]. Combining the probabilistic notion of controllability with the probability of the hazardous event itself occurring, practitioners are able to classify risk from a more manageable set of probability / severity combinations.

The Controllability of Automotive Safety Targets (CAST) project [114] used vehicle simulation to validate the assumptions made by the DRIVE project about the inverse relationship that exists between controllability and the probability of having an accident following a hazardous event. Additionally, for the vehicle hazards simulated, the study was able to correlate the simulation results with a desk based assessment of controllability completed beforehand.

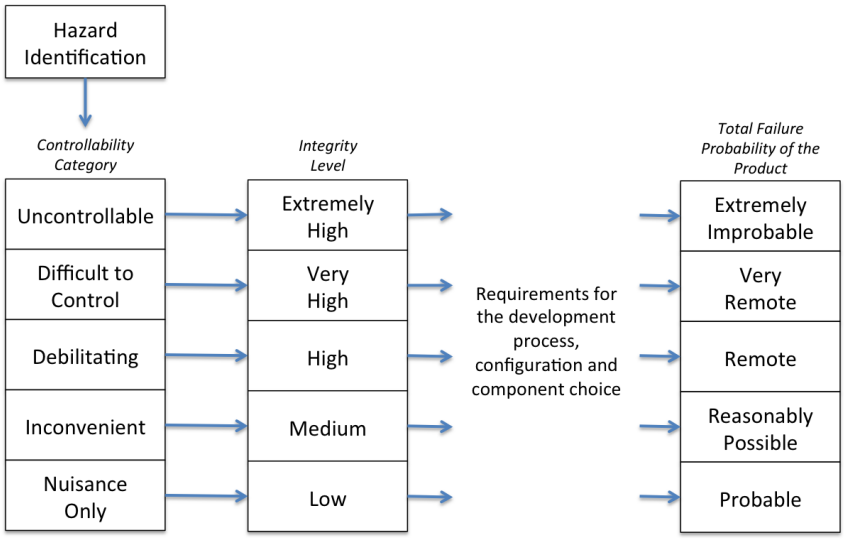


Figure 10: Controllability and integrity categories from DRIVE (reproduced from [85])

6.2 AUTOMOTIVE MODEL OF DRIVING

As briefly discussed in Section 1.2.2 (page 17), the automotive functional safety approach makes implicit assumptions about the driver and the vehicle. These assumptions undoubtably influence controllability assessments. Although inconsequential in a fully manual driving context, adding automation does potentially bring these implicit assumptions into question.

The first implicit assumption is that the driver is always fully in control and responsible for their vehicle's behaviour in traffic. This assumption originates from the 'Vienna Convention' (see Section 1.2.2) [178]. The Vienna Convention is an international treaty which by establishing 'standard traffic rules' for such things as vehicles, road signage and driver obligations, seeks to facilitate cross-border traffic movement and improve road safety.

The second assumption is that the driver is integral to vehicle control, as illustrated by the MISRA VCM (see Figure 1 page 19). Being in complete control

of the vehicle 100% of the time, the implication is that the driver is, and remains, fully aware of their surroundings. With this situational awareness encompassing such things as: the vehicle behaviour, the behaviour of other road users, and changes in vehicle response or performance (perhaps due to changing environmental conditions or vehicle system failure).

The third assumption relates to how vehicle technology has evolved over the last 100 years or so, and the gradual introduction of E/E programmable control systems. Typically automotive systems have been engineered with a *fail passive* safe state[1]. That is, the detection of a failure within a vehicle system leads to that function being shut-down. The MISRA "state-machine model of automotive risk" (Figure 11) illustrates this implicit relationship between vehicle control, controllability and system failure [113].

6.2.1 *Exploring controllability with familiar automotive examples*

To illustrate how the notion of controllability might be used with the MISRA VCM (see Figure 1, page 19), the following vehicle features are discussed: an engine control system, an air suspension system, and an electric driveline incorporating in-wheel motors.

As introduced in Section 1.2, emissions legislation, fuel economy, performance and cost all contributed to engine control systems evolving from a purely mechanical system to a complex programmable control system. Although the earliest mechanical systems had the potential to cause *Unintended Acceleration* (for example, by the accelerator cable becoming stuck) the introduction of electronic throttle control systems led the hazard, and its associated risk classification, to come into focus.

Although the early engine control systems typically mimicked simple control relationships, previously achieved mechanically, modern engine control systems utilise complex control algorithms expressed in the torque domain.

1 ISO 26262 defines a safe state as the "operating mode, in case of a failure, of an item without an unreasonable level of risk" [162].

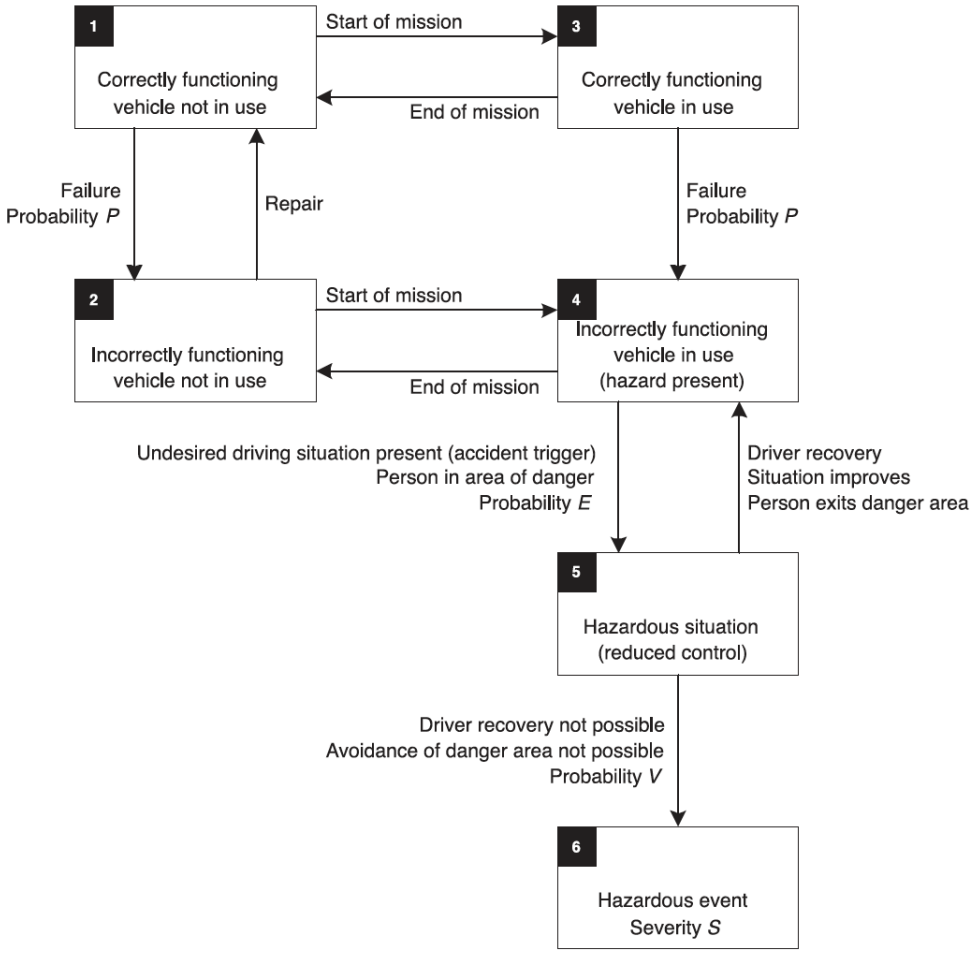


Figure 11: MISRA State-Machine Model of Automotive Risk (reproduced from [113])

From a systems safety perspective this leads to a complex relationship between the driver's accelerator pedal position and the resultant vehicle acceleration. Typically layered monitoring strategies [43] are used to continuously monitor the relationship between the driver's torque demand (via the accelerator pedal) and the estimation of delivered engine torque (estimated from measured engine parameters); with any hazardous discrepancy between demanded and delivered torque being mitigated by limiting the engine's performance or ultimately by shutting the engine down.

The above engine control torque-monitoring concept fits seamlessly with the notion of controllability, with the implicit assumptions made about the driver (described above). For an experienced driver the longitudinal control task is generally a subconscious one that can be described using the elements

of the MISRA VCM as follows: The driver modulating the accelerator pedal position (*Driver Control*), based on their current *Driver Strategy* and given their knowledge of current *Traffic Environment*, to achieve the desired vehicle acceleration. If the vehicle's acceleration (*Complete Vehicle Behaviour*) is too slow then the accelerator pedal can be pressed harder. Conversely, if vehicle acceleration is too great then the accelerator pedal can be released, and once completely released the brakes applied.

To mitigate the hazardous effects of engine control system failures, controllable acceleration targets can be defined, e.g. *"Vehicle positive longitudinal acceleration shall not exceed driver demand by* $> 1.5\ \mathrm{ms}^{-2}$ *for longer than* $1\mathrm{s}$*"*, knowing the driver's likely response [24]. The mitigating action *"outputs are electronically 'limited' to a fixed value"* can then be defined, thus ensuring that any engine torque anomalies detected are prevented from exceeding a level where the resultant acceleration might be difficult to control. With these targets defined the torque monitor's internal parameters can then be set to achieve the required response.

Air suspension is another example where the notion of controllability works well. Air suspension systems of the type used on large luxury four-wheel drive sport utility vehicles control the height of the vehicle and allow changes to be made to the vehicle's ride height; with ride height being raised to facilitate off-road driving and lowered to make vehicle ingress and loading easier [66]. As selecting an incorrect ride height could affect the vehicle's handling characteristics (changing the relationship between *Driver Control* and *Complete Vehicle Behaviour* in Figure 1 page 19) vehicle speed limits constrain when ride height changes can be made. Consequently a failure leading to the vehicle being at the wrong ride height, for a given vehicle speed, may result in a hazard.

Typically safety mechanisms detect failures within the air suspension system that could lead to an incorrect ride height selection. Having detected a failure, and dependent on the current *Vehicle Manoeuvre*, the safety mechanisms can then either disable the system (*Fail Passive*) and issue driver warnings (dashboard warning lights, messages or audible chimes) or disable the system

and limit the speed. In either case the failure mitigation makes assumptions about the driver's place within the overall control loop: In the first case, once the driver has been warned about the potential changes to vehicle handling they will change their *Driver Control* accordingly. In the second case, where vehicle speed is limited, the vehicle's operating envelope (or choice of *Vehicle Manoeuvres*) is constrained into the region where the vehicle is known to be controllable.

Even for the development of novel technologies, such as an in-wheel motor electric driveline system [44] the notion of controllability still works because the assumption that the driver is a part of the control loop is still true.

Although the ability to control torque delivery to each road wheel independently is a vehicle dynamists dream, and a major selling point for in-wheel motors, applied incorrectly this torque asymmetry causes the hazard *Induced Yaw*. Empirical driver response studies [118] set yaw rate and lateral acceleration controllability limits from which a power asymmetry target of 30 kW / tonne could be derived [45]. Once set, safety mechanisms then monitor the asymmetric power being produced by a pair of motors, and either shut-down or modify motor control if the power asymmetry exceeds the limit defined. Consequently, system failures having the potential to adversely affect the *Vehicle Control System* can be detected and *Vehicle System Behaviour* influenced such that the *Complete Vehicle Behaviour* remains controllable by the driver.

6.3 THE IMPACT OF GREATER AUTOMATION

6.3.1 *Pivotal literature*

Written by Bainbridge in 1983, but still highly relevant today, could "Ironies of automation" [17] provide insight into how the driver's relationship with the car will change with greater autonomy? And could this then give further insight

into how engineering assumptions made during design may need to change to ensure that safety is maintained?

In the paper Bainbridge discussed how greater automation in an industrial process context might actually increase rather than decrease human operator problems, particularly if greater automation changes the human's role to that of the supervisor required to intervene under abnormal conditions. This relationship between the human operator and automation is discussed through a number of *'ironies'*, of which a number are discussed below.

A second paper potentially holding more clues is "Humans and automation: Use, misuse, disuse, abuse" [128], written by Parasuraman and Riley 14 years after Bainbridge published her work. Again ideas presented by these authors are considered in the context of vehicle automation below.

6.3.2 *The use of automation*

As automation increases and vehicle systems gain greater authority, it is perhaps logical to expect that *Driver Control* will become a shared task, between the human driver and the AD system, but what else could these pivotal papers tell us about the notion of controllability in an highly automated driving vehicle context?

Clearly one motivation for the introduction of AD vehicle systems is the reduction in accidents resulting from human error in the *Driver Control* task. Perhaps the first challenge is correctly choosing which tasks to automate, as deploying automation when it is not really the right option is one abuse of automation, potentially leading to increased driver workload [128].

Automation has supported the driver in the longitudinal vehicle control task for many years. The earliest cruise control systems simply controlled vehicle speed to a pre-set value, enabling the driver to simply set a constant vehicle speed while manually controlling the remainder of the *Driver Control* task. Therefore providing that the vehicle is in a *Physical Environment* where the driver is afforded sufficient space and time in which to react, cruise control

failures would normally be considered easily controllable. One way to ensure the driver does have sufficient space and time in which to react is to constrain the cruise control's ODD. Generally this is achieved by enforcing a minimum set speed below which the feature is disabled.

ACC adds functionality to standard cruise control, thus supporting the *Driver Control* task further. Although while active, the longitudinal control aspects of *Driver Control* are fulfilled by the ACC system, the driver is not dissolved of all responsibility for the task. In some operational situations, such as the preceding vehicle braking sharply, the ACC may require the driver to resume manual control. Generally this occurs because the ACC system does not have sufficient authority over the vehicle brake pressure needed to achieve a full emergency stop. Therefore while ACC is active the driver inherits a new monitoring task, the consequences of which are that the driver must understand ACC operation sufficiently to know when manual control may be required [128], and remain vigilant and ready to intervene when required [17].

A vehicle fitted with an ADAS, such as ACC, also places new responsibility on the driver, mainly knowing when to give longitudinal control to the ACC system and when to undertake *Driver Control* manually. Automation misuse can occur when users become over reliant on automation [128]. For example attempting to use ACC when environmental factors (such as being in fog or on very windy roads) limits system performance or leads to unexpected system behaviour and potentially system disuse because the driver's confidence in the system is lost [128].

Leaving the decision about when and where to enable automation to the driver's discretion potentially leads to unpredictability in the use of automation, and it is suggested that different people use different strategies when choosing whether to use automation [128]. Potentially this could add complexity and uncertainty into the design task. Additionally, gathering the verification evidence that enables the design team to demonstrate system safety potentially becomes a far bigger task. This is because the MISRA VCM (Figure 1 page 19) effectively becomes multi-dimensional; with a separate dimension existing for each combination of the *Driver Control* task.

6.3.3 *The driver's task with automation*

As previously discussed one intuitively knows that the driver's role will change, but what might the new role become and will the driver be more or less effective in that new role?

ACC highlighted how the driver's role may become a monitoring task. The irony here is that humans are less good at monitoring automated tasks than they are at carrying out the task manually [17]. Also if the manual task was largely achieved subconsciously then automating that task and asking the driver to monitor the automation could actually add to the driver's workload [128]. LKA illustrates where automation supports a subconscious task. Steering is a good example of a *Driver Control* task requiring little cognition by the driver. It is a task that the driver practices constantly during each journey, so it quickly becomes completely subconscious even for the most novice driver. LKA is designed to support lateral vehicle control by applying a corrective torque (either through the hand wheel or by an asymmetric brake application), if the driver becomes distracted and allows the vehicle to stray from its lane.

The automation of a task of this type raises questions about the actions that should be taken in the presence of failure. For systems like LKA, failing *passively* may not be a viable option. And what about the controllability assumptions that would have been made regarding other systems fitted to the vehicle? For example, the power assisted steering system and the hazard loss of steering assist? As the introduction of LKA effectively adds a use case, will the assumptions made previously still hold true? The implications of this are perhaps exacerbated by the fact that a driver, needing support to maintain their position in lane, is probably indicative of the driver who has lost situational awareness! The loss of situational awareness is known to adversely affect driver reaction times [106], which may further complicate the choice of mitigation.

6.3.4 *The designer's task with automation*

Both papers [17, 128] discuss the importance of the design team and in particular their view of the human operator. If ADAS features are developed with the sole motivation of removing the unreliable and inefficient driver from the control loop then Bainbridge warns of two potential outcomes [17]: motivated by the desire to remove the human from the system may actually lead design teams to target the easy to automate *Driver Control* tasks first. The side effect of which is to leave the driver with the more difficult *Driver Control* tasks still to be performed manually.

An inevitable consequence of greater automation is also increased complexity and as highlighted above the MISRA VCM (Figure 1, page 19) may actually become a multi-dimensional problem; with the potential for subtle interactions between the dimensions. The desire to remove human error may simply move the errors to the design task, as "one cannot remove the human error from a system simply by removing the human operator" [128]. Today's vehicle is a complex system of systems having the potential to exhibit both good and bad emergent properties; so perhaps the driver's contribution to *Driver Control* is more crucial than ever before [17].

6.4 SUMMARY

The current approach to automotive system safety assumes that the driver is an integral part of the vehicle control loop and, like many other standards, ISO 26262's risk model includes the notion of controllability.

Two pivotal human factors papers have been used to 'test' the notion of controllability and the MISRA VCM (Figure 1 page 17). Although these models work when there is no automation, even for SAE Level 1 ADAS (like ACC) deficiencies become apparent. The most notable omission is perhaps the lack of feedback from the *Vehicle System Behaviour* to the *Driver Control* or *Driver Strategy*.

The notion of controllability has clearly been used to good effect in the past, particularly with regard to managing the complex automotive hazardous event space. However, introducing greater automation potentially invalidates the risk model – with greater automation gradually removing the driver from the vehicle control loop, and vehicle connectivity making the hazardous event space multi-dimensional. This drives the need for the risk model to be re-evaluated and the term controllability precisely defined. The MISRA VCM (Figure 1 page 19) is indeed useful, but it does need to change to consider the multi-dimensional nature of the problem.

Part III

RESEARCH CONTRIBUTIONS

Part III of this thesis describes the products created by this research. Before describing each research product in detail, Chapter 7 first describes how the three products combine to form a hazard analysis process for shared control.

The first research product is a conceptual model referred to here as the Enhanced Vehicle Control Model (EVCM). By including elements representing the shared cognitive nature of the driving task, the MISRA VCM has been reimagined to extend its utility to AD vehicle features. When used in conjunction with the accompanying shared control hazard analysis method (the second research product), the EVCM helps conceptualise the nature of AD shared control to facilitate the identification of hazard causes. This contributes a new dimension to automotive hazard analysis not covered by the current automotive safety standards [162, 165] and extends the 'state of art'. The third research product is a safety case argument pattern for the shared control aspects of the AD feature's development. The intention is that this safety case argument pattern will form part of the full safety case argument for the AD feature.

7

RESEARCH CONTRIBUTION OVERVIEW

The preceding chapters have sought to describe the increasing prominence that automation has in passenger vehicles today and the motivation for that proliferation. As discussed in Chapter 2 vehicle automation is having an increased authority over the DDT. This changes the driver's role, which directly affects the assumptions that can be made about the driver and their ability to maintain control should an automation fault occur.

The established automotive approach to functional safety relies on the driver. As discussed in Section 1.2.2 this functional safety paradigm assumes that the driver is integral to vehicle control; a consequence of which is that the driver remains situationally aware. This notion that the driver is at the heart of vehicle control has been the foundation of the ISO 26262 automotive risk model. With the risk classification schema including a *controllability* term (see Chapter 6), which allows credit to be taken for the driver's ability to maintain vehicle control following a hazard and avoid an accident.

In Chapter 3 the case study involving practising automotive functional safety practitioners and academics brought into question the continued utility of the existing hazard analysis and risk classification method. In particular, the need to consider differently (and perhaps even redefine) the notion of controllability, and to broaden the scope of hazard analysis from a single vehicle-centric view of system functionality.

Part III of the thesis introduces a hazard analysis method for shared control. This is shown pictorially in Figure 12. To orientate the reader an overview of the process is introduced in this chapter. The subsequent chapters within Part III provide more detail about: the EVCM development, the use of behavioural competencies to describe AD feature functionality and to define

agent responsibility, the accompanying hazard analysis method, and the safety case argument pattern for shared control.

7.1 AD FEATURE BEHAVIOUR AND AGENT RESPONSIBILITY

Fundamental to the analysis of shared control is the EVCM which is described in detail Chapter 8. The EVCM allows the shared nature of AD vehicle features to be conceptualised at an appropriate level of abstraction to allow potential shared control hazard causes to be fully explored. On its own the EVCM has no mechanism by which to describe AD feature functionality. However, when used in conjunction with the behavioural competency taxonomy (see Appendix A), the AD feature's behaviour can also be described.

7.2 SHARED CONTROL HAZARD ANALYSIS METHOD

Chapter 9 describes using the EVCM in an STPA based analysis method to uncover loss scenarios attributable to an AD feature's shared control. Throughout Part III reference is made to *Classic* STPA. This is the STPA method described within the *STPA Handbook* [96], and that forms the foundation of the *Shared Control* STPA method (see Figure 12).

Like *Classic* STPA the *Shared Control* STPA method starts by identifying the stakeholders, losses and hazards attributable to the AD feature (see ① in Figure 12). An additional prerequisite activity then defines the responsibility split between the automation and the driver - also referred to as the machine and human agents. Identifying the behavioural competencies needed by the automation to achieve the intended functionality, but also the behavioural competencies that the human agent needs to be performed, to maintain safety within the ODD, informs the scope of shared control. This knowledge is then taken forwards into ②.

Step 2 (see ② in Figure 12) populates a version of the EVCM drawn in a hierarchical control structure style. Using the knowledge gained from considering

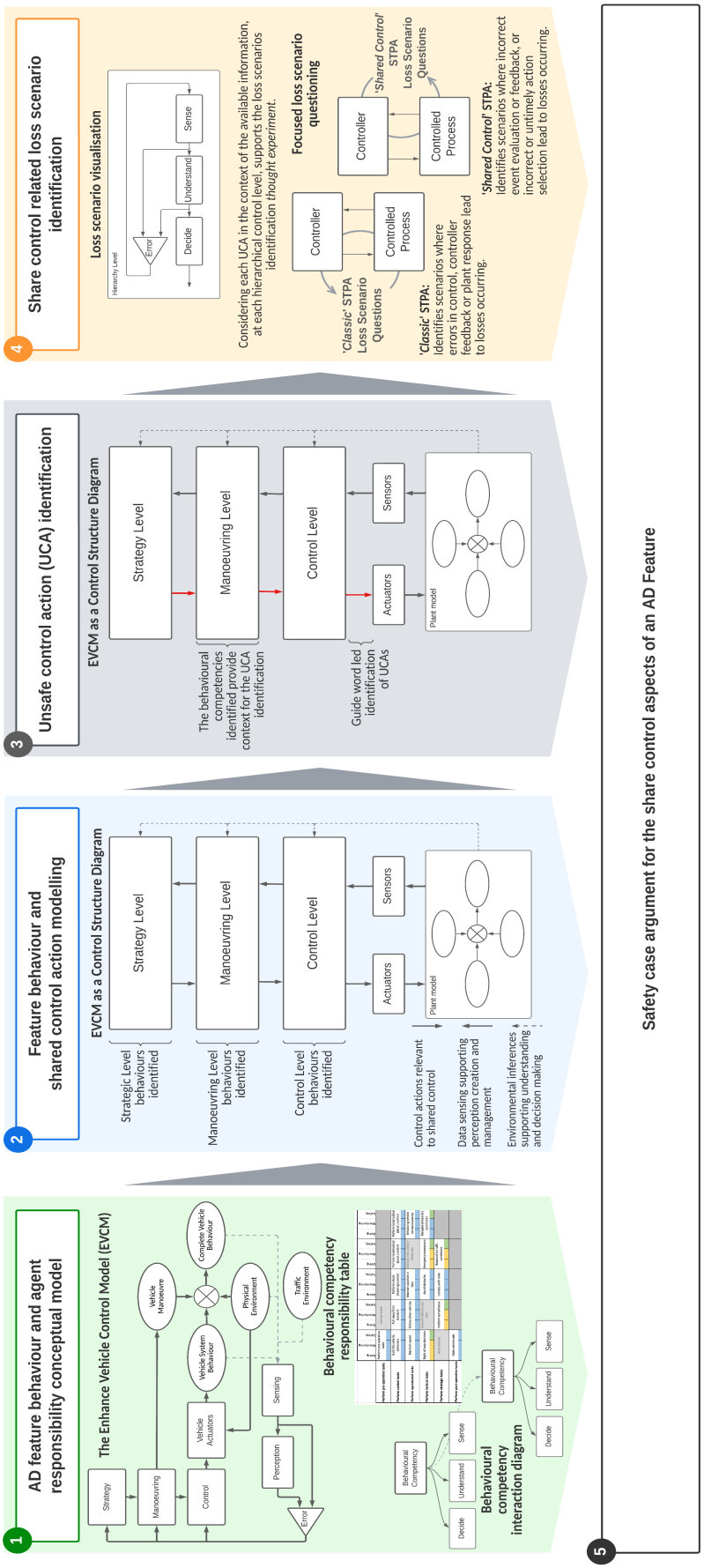


Figure 12: The shared control hazard analysis and safety case argument process

the shared control responsibility split to populate the CSD. Specifically, the pertinent control actions, the sensed data, the information inferred from the sensory data, and the other inferences made about the external environment that should also be included in the analysis.

Like *Classic* STPA, *Shared Control* STPA step 3 in the process (see ③ in Figure 12) uses a systematic guide word lead process to assist the analyst to identify the UCAs. However, the *Shared Control* method enhances the UCA identification process by using the behavioural competencies identified in ① to uncover further UCAs.

The final step (see ④ in Figure 12) identifies the scenarios in which the identified UCAs will lead to hazards, and hence losses. To address shared control hazard causes, two further loss scenario types are introduced. Guided questioning and a SUDA visualisation of the decision making cycle facilitates the loss scenario identification *thought experiment*.

7.3 A SAFETY CASE ARGUMENT FOR SHARED CONTROL

Unlike other transport sectors the use of safety cases incorporating structured safety case arguments are less prominent in the automotive industry. However, this is beginning to change. ISO 26262 Edition 2 [162] now explicitly calls for a safety case argument to be documented. In addition, the SOTIF standard ISO 21448 [165] discusses the importance of documenting a SOTIF safety case argument that complements the functional safety activities [165, Clause A.2.3].

Demonstrating that hazard causes, resulting from the shared nature of control, have been identified and correctly addressed is an important aspect of AD feature safety. The safety case argument pattern (see ⑤ in Figure 12) presented in Chapter 10 builds such a safety case argument for shared control, using the evidence created by following the *Shared Control* hazard analysis method described in Chapter 9.

8

A CONCEPTUAL MODEL OF SHARED CONTROL

8.1 INTRODUCTION

As the previous controllability discussion (see Chapter 6 page 94) demonstrates the MISRA VCM (Figure 1 page 19) remains valid for vehicle systems where the Driver is integral to vehicle control and has utility supporting hazard analyses [109]. However, once a part of the *Driver Control* element is replaced with automation, then model deficiencies become apparent. The following section describes the development of an Enhanced Vehicle Control Model (EVCM). A conceptual model that remains relevant and useful for the analysis of highly automated and connected vehicles.

8.2 EVCM CONSTRUCTION

Initial attempts to modify the MISRA VCM involved splitting the *Driver Control* element between the human and machine actors, and using data-information-knowledge-wisdom pyramids[1] [16] to highlight the different information used by human and machine perception. Having taken this approach, two things quickly became apparent: Firstly, this approach simply added complexity to the diagram, which intuitively would never lead to an elegant solution. Secondly, it did highlight the stark contrast between how the human perceives the environment and how the automation would perceive the same environment. This raised the question as to whether this difference could be used to

1 Data-information-knowledge-wisdom pyramids were introduced during the data inference discussion in Section 2.3.4.2 on page 40.

highlight potential safety concerns, particularly when control is being shared or transferred between the human and the automation.

The literature pertaining to driver behaviour and modelling suggested Michon's HCM [107] (see Figure 2 page 27) as a useful model for describing *Driving Control* from the human driver's perspective (referred to herein interchangeably as the *human agent*). However, to be useful in the context of automated driving, the enhanced model must also facilitate modelling functions carried out by the automation (referred to herein interchangeably as the *machine agent*). Therefore, the question becomes "can automation functions also fit with the three levels of the HCM?"

In their discussion about multi-sensor environmental perception Schubert and Obst present a perception layer interfacing model (see Figure 13 diagram (a)). This they used to describe the interface between sensors and the higher level automated driving functions [137]. Considering the three models side-by-side (see Figure 13 diagrams (a), (b) and (c)), a relationship is drawn between the three models, making the enhanced representation of the *Driver Control* element possible.

While the meaning that one should infer from each MISRA VCM element might be intuitive in a manual driving context, the meaning becomes less clear in a highly automated and connected vehicle context. This necessitates defining the underlying meaning of each model element (see Section 8.3). The MISRA VCM also used solid and dotted lines without explanation (see Figure 1 page 19). This research interprets the solid lines to represent physical interfaces or interactions, and the dotted lines to represent the perception, interpretation or understanding of something physical. Reviewing the model resulted in the inclusion of a mechanism that represents updating agent prior knowledge (mental models), and future state predictions, based on sensory perception [48]; effectively, the notion of obtaining and maintaining SA (see Section 2.1.2 page 27). This is represented in the enhanced model by a comparator. If the error between a given mental model and what is being perceived becomes

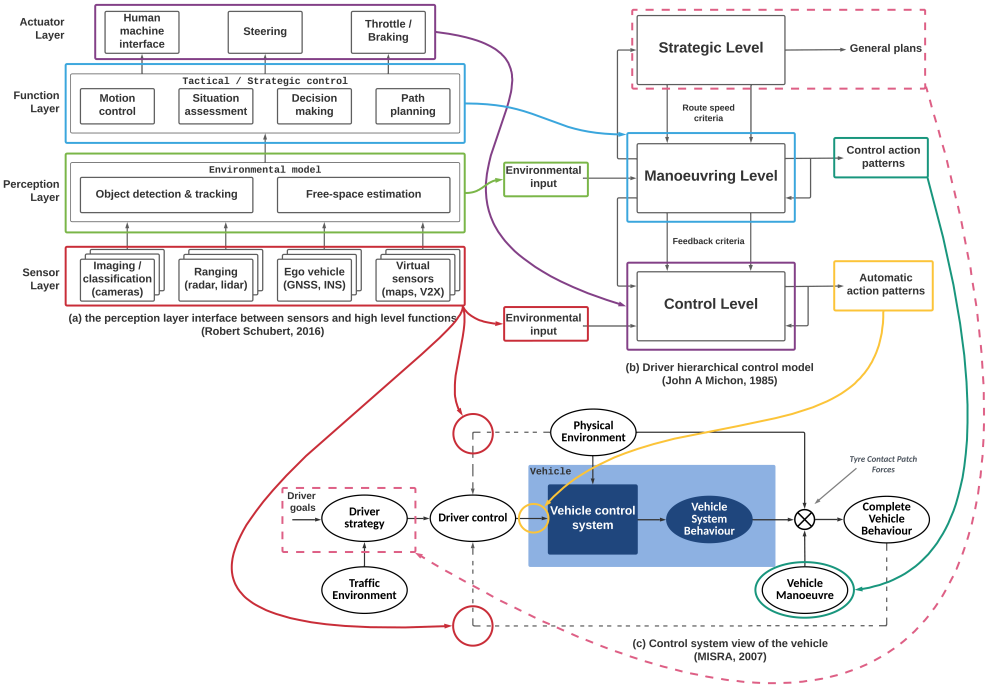


Figure 13: The relationship between the three models that combine to form the EVCM (reproduced from [107, 113, 137])

unacceptably large, then cognitive awareness moves up through the hierarchy of control [C. McCall[2], personal communications, 6th December 2017].

Combining these concepts gives rise to the Enhanced Vehicle Control Model (EVCM) (Figure 14).

8.3 ELEMENTS OF THE MODEL EXPLAINED

8.3.1 *Strategy level*

Navigating the EVCM (Figure 14) from the top left, the first input is *Driver Goals*, which are the driver's[3] high-level objectives (e.g. drive to work, take the kids to school, or just go for a ride in the country). Achieving such goals

2 Dr. Cade McCall is a Lecturer and Assistant Professor in the Department of Psychology at the University of York. As an experimental psychologist he uses virtual environments to study emotion, cognition and behaviour in threatening scenarios.

3 Or vehicle occupants in a fully automated driving context.

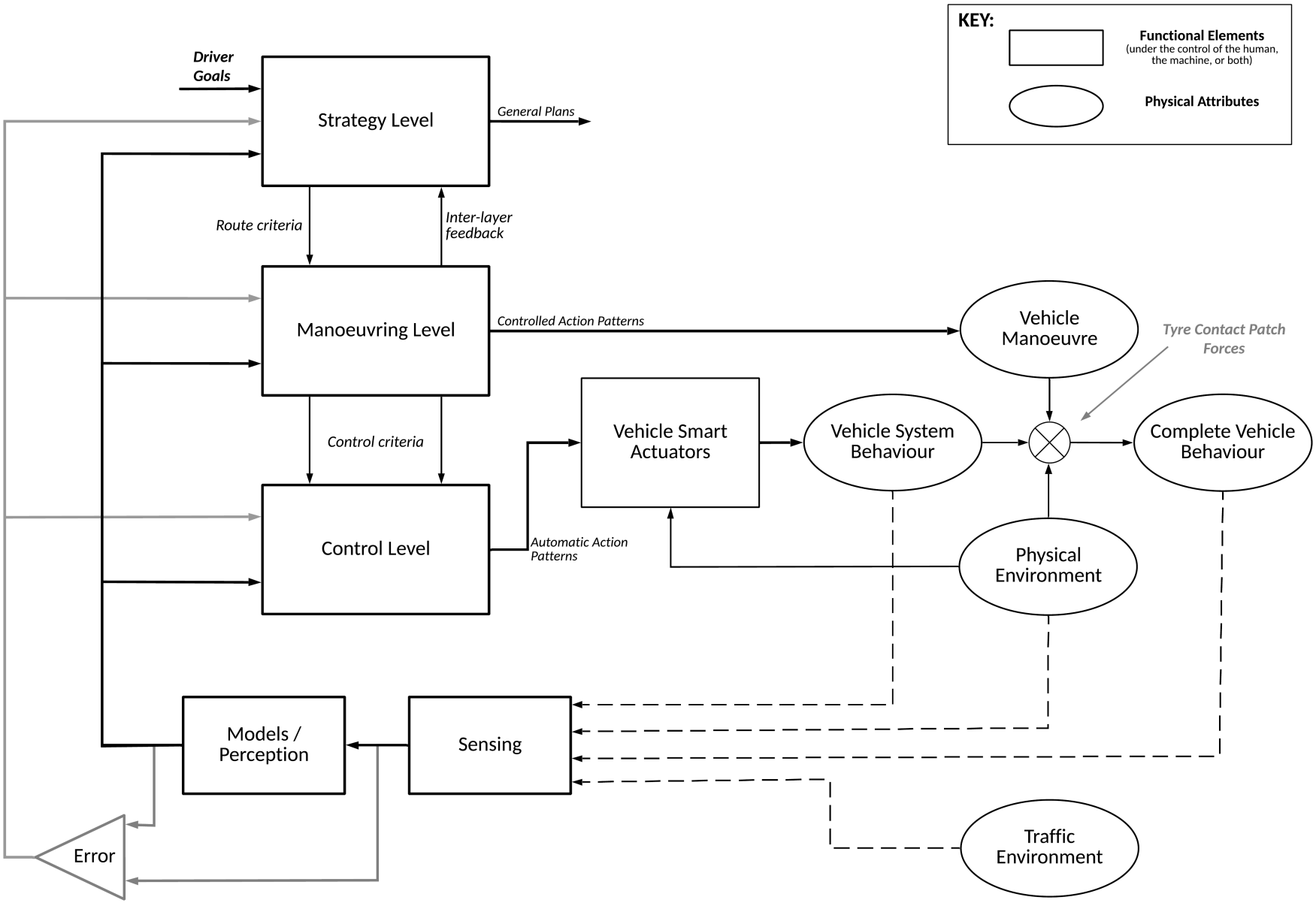


Figure 14: The Enhanced Vehicle Control Model (EVCM)

requires strategic planning, represented by the *Strategy Level* box. Applicable behavioural competencies for the *Strategy Level* (see Section A.1 page 211) include: route planning, performing surveillance, and responding to traffic conditions. Such strategic type routing and planning decision making will be influenced by external information (e.g. the current *Traffic Environment*), any goal related constraints (e.g. to arrive by a particular time), and progress related feedback received from the *Manoeuvring Level*. Typically strategy level tasks occur over a long time-frame and do not require a real-time response [49]. *Route criteria* represent the targets and constraints that the *Strategy Level* places on the *Manoeuvring Level* below. For example, the level of urgency associated with a given journey; that is, a leisurely drive becoming stressful as heavy traffic adversely affects progress.

8.3.2 *Manoeuvring level*

Moving anti-clockwise around the model, the *Manoeuvring Level* represents the tactical and operational driving behavioural competencies (see Section A.2 page 212), like negotiating a junction or determining who has the right-of-way during a particular manoeuvre. These behavioural competencies will be rule-based real-time tasks that typically last a few seconds [49]. They are predominately data driven, largely constrained by the characteristics of the given situation, and are in support of the *Route criteria* goals and targets set by the *Strategy Level*.

In Michon's HCM (Figure 2 page 27) the *Manoeuvring Level* has three outputs, *Controlled Action Patterns*, *Feedback Criteria*, and an unnamed path to the *Control Level*, although the literature does not convey the intention of these outputs. Here *Controlled Action Patterns* are taken to represent the rule-based [131] action sequences needed to carry out the intended manoeuvre, while *Feedback Criteria* are called *Control criteria* and represent the information needed by the *Control Level* to successfully complete the chosen manoeuvre. For example, knowing that on a cold day the road might be slippery so choosing to limit the magnitude of steering and braking inputs. The intention of the third output from the

Control Level is unclear. However, given that it appears to flow from the *Control Action Patterns*, through the *Manoeuvring Level* and out to the *Control Level*, the hypothesis is that it is a feed-forward control path for the level below. For example, beginning to apply a steering input prior to the vehicle entering the corner. The EVCM includes the notion of feed-forward control within the *Control criteria* output.

In addition to three outputs, the *Manoeuvring Level* in Michon's original model has three inputs – one from the environment, one from the *Strategy Level* (discussed above), and one a feedback path from the *Controlled Action Patterns*. The EVCM maintains a similar structure. Like the HCM (see Figure 2 page 27), the *Manoeuvring Layer* receives environmental information, but via the *Sensing* and *Models / Perception* path rather than directly. Feedback from the *Control action patterns* also comes via the *Sensing* and *Models / Perception* path, rather than directly as is the case in the HCM. In both cases, the information received via the *Sensing*, *Models / Perception* and *Error* feedback paths will affect the *Manoeuvring Level's* subsequent actions, and any errors identified will likely influence future perception, understanding and decisions making.

8.3.3 *Control level*

Below the *Manoeuvring Level* is the *Control Level*, which contains the low level driving behaviours that are largely skill-based [131], automatic and occurring in the millisecond time-frame [49]. Example *Control Level* behavioural competencies include controlling the vehicle's steering, acceleration and braking (see Section A.3 page 217). Being the basal level in the hierarchy, the *Control Level* has only one output – the *Automatic Action Patterns*. This output comprises the low-level control actions to the vehicle control system, referred to as the *Vehicle Smart Actuators* within the EVCM.

Like the other levels, the *Control Level* has inputs capable of modifying its behaviour. In Michon's HCM these were the environment, *Automatic Action Patterns* feedback and the *Feedback Criteria* from the *Manoeuvring Level*. In

the EVCM the environmental inputs, both about the physical environment (e.g. carriageway edge, air temperature) the *Complete Vehicle Behaviour* (e.g. lateral and longitudinal acceleration), and *Automatic Action Patterns* feedback (e.g. steering torque, display information) all come via the *Sensing* and *Models / Perception* blocks.

8.3.4 *Sensing*

Towards the bottom of the model, the *Sensing* block represents both the human senses and the sensors used by the automation to collect *Vehicle System Behaviour*, *Complete Vehicle Behaviour*, *Physical Environment*, and *Traffic Environment* data. For the human this might be sight and the vestibular system, for the automation this might be RADAR, LiDAR and cameras.

8.3.5 *Models, perception and errors*

The *Models / Perception* block is included to represent the idea of the controlling agent (i.e. human, automation, or both) obtaining and maintaining SA within the model. The *Model / Perception* block represents both the model of the environment in which the vehicle is operating, and the perception of the vehicle's behaviour within that environment. For example, the model might include information about stationary objects in the vehicle's physical environment or data regarding the current vehicle speed. These models will be developed from information inferred from sensory data via the *Sensing* block, and updated accordingly as new data is received.

From a control perspective the path from physical system attributes (*Vehicle System Behaviour, Complete System Behaviour, Physical Environment, Traffic Environment*) through *Sensing* and *Models / Perception* can act as a control path for both feedback (or compensatory) control or feed forward control [49]. Large errors between the output of the *Models / Perception* and new data being received via the *Sensing* block will potentially influence the *Control, Manoeuvring* and

Strategy level blocks and drive changes in decision making. The threshold at which this change in decision making occurs might affect safety. For example, the human might take back control prematurely because they perceive an issue occurring, or they may continue to use the automated system inappropriately because they are unaware of inbuilt limitations [183]. The EVCM incorporates this idea by including an *Error* block having feedback paths to the *Control Level*, *Manoeuvring Level* and *Strategy Level* blocks.

The *Error* block can also model the stark differences in environmental perception capability that may exist between the human and the automation, potentially leading to unexpected or uncoordinated interactions between the automation and the human driver. For example, if one considers the behavioural competency of following other vehicles, in the context of ACC, the potential effect of perception differences for the cut-in scenario become evident. In this scenario the automation is unable to perceive a vehicle cutting in from an adjacent lane, so will not brake until the 'cut in' vehicle is well into the *ego vehicle's* own lane. In contrast, with the driver still in the control loop and situationally aware, they will typically react quickly to other visual cues and cancel ACC, well before the other vehicle encroaches into their lane.

8.3.6 *Vehicle plant model*

The last part of the EVCM is the plant model. The vehicle plant model comprises the *Tyre Contact Patch* and the physical system attributes that influence and are influenced by the *Tyre Contact Patch*. This remains largely unchanged from the original model (Figure 1 page 19). The *Tyre Contact Patch* is the small area where the tyre and road surface are in physical contact, with an amount of slip generating the friction needed for the tyre to apply a force to the road surface, either laterally, longitudinally or both. The physical system attributes of *Physical Environment*, *Vehicle System Behaviour* and *Vehicle Manoeuvre* all influencing the forces being applied to the road by the tyre, while the *Complete Vehicle Behaviour* is influenced by the forces that result from the tyre-road interaction.

The *Vehicle Manoeuvre* represents the manoeuvre currently being carried out by the vehicle and will influence how the *Vehicle System Behaviour*, applied at the *Tyre Contact Patch*, translates to the *Complete Vehicle Behaviour*. For example, entering a corner too quickly may lead to excessive side-slip resulting in vehicle under-steer. The *Vehicle System Behaviour* is the vehicle level behaviour that occurs in response to system level changes made by the *Vehicle Smart Actuators*. For example, positive or negative torque being applied to the vehicle's driven wheels. Next, the *Physical Environment* represents the environment in which the vehicle is being operated (e.g. weather). This will influence how the *Vehicle System Behaviour* translates to *Complete Vehicle Behaviour*, but also environmental characteristics like temperature will also influence the *Vehicle System Behaviour*. Finally *Complete Vehicle Behaviour* represents the physical vehicle response, which can be described using the 6 degrees of freedom rigid body model [61].

8.4 DESCRIBING FEATURE BEHAVIOUR

As stated at the beginning of Section 8.2, while it is conceivable that in a manual driving context correct behavioural inferences will be made about model elements, the same is not true in an automated driving context. Here it is important that the AD feature behaviours needed to achieve the intended functionality are explicitly stated. For example, maintain a constant vehicle speed or maintaining the desired *headway* to the preceding vehicle in the case of ACC. In addition, the behaviours needed to achieve and maintain vehicle safety within the ODD should also be explicitly stated. This activity of identifying the behaviours needed to achieve and maintain vehicle safety also has the effect of highlighting where shared control exists. This information is vital to the *Shared Control* STPA method described in Chapter 9.

The behavioural competency taxonomy (first introduced in Section 2.1.3) provides the language from which to describe an AD feature's shared control. Behavioural competencies not applicable to the ODD can be discarded, while behavioural competencies that the AD feature automation needs to deliver or

that are needed to maintain vehicle safety can be identified. Having identified all relevant behavioural competencies, the responsibility[4] for undertaking those behavioural competencies can be determined. By identifying whether it is the human, the automation, or both that are responsible for a behavioural competency helps define the nature of shared control for the AD feature.

An example behavioural competency responsibility table appears in Section 9.3.1 Table 4 on page 128. Table 4 has been completed for a partial automation vehicle feature called HAS that influences both longitudinal and lateral vehicle motion.

Using a similar approach to the STOE-KRS framework (first introduced in Section 2.3.3), each behavioural competency is considered in turn, and the decision made as to whether the competency is relevant to the AD feature and its ODD. Of those behavioural competencies deemed relevant, then the decision is made about which agent is responsible for undertaking the competency (the *doing*), which agent is responsible for monitoring task execution (the *monitoring*), and finally which agent has the responsibility for maintaining vehicle safety (the *achieves safety*). In this context the agent is either the human driver or the automation, and the responsibility for each competency can be either the human's, the automation's or it can be shared. Once completed, not only does the behavioural competency table provide a comprehensive definition of the AD feature's intended functionality, but also highlights the nature of shared control.

Together the EVCM and the completed behavioural competency table for an AD feature provide the analyst with the prerequisite material for the *Shared Control* hazard analysis method described in Chapter 9.

4 Responsibility in this context refers solely to role-responsibility. That is, S has a specific duty to bring about X as part of an assigned task or by agreement. Other responsibility classes, such as causal, legal, moral and capability that those readers familiar with Hart's work may be aware of, are not considered here [71].

8.5 SUMMARY

The origins and construction of the EVCM has been described. Given the potential complexity of any AD feature HMI being modelled, it is important to remove any ambiguity in meaning from the model itself. This has been addressed by describing each model element and information flow in turn. The behavioural competency taxonomy has then been introduced as the means by which the AD feature behaviour can be described. Of particular importance to AD shared control safety is making the Human Machine Interactions between the human and machine agents explicit. This is achieved by completing the behavioural competency responsibility matrix for the AD feature in question. That is, clearly stating the agent who is responsible for undertaking the task (the *doing*), the agent who is responsible for monitoring the task while underway (the *monitoring*), and the agent who is responsible for maintaining vehicle safety (*achieves safety*).

The EVCM, in the form shown in Figure 14, can be used to model an AD feature. However, as will be shown in Chapter 9, the EVCM can be redrawn as a hierarchical CSD to facilitate STPA based hazard analysis.

9

SHARED CONTROL HAZARD ANALYSIS METHOD

9.1 INTRODUCTION

Having described the EVCM (in Chapter 8), this chapter describes an accompanying hazard analysis method. The analysis method builds on the STPA method, adding cognitive considerations applicable to shared control. Used together, the EVCM and the *Shared Control* STPA method allows the hazards and hazard causes associated with shared control to be investigated.

To describe the steps that comprise the method an AD vehicle feature is used as a running example. This chapter first describes that AD vehicle feature before then describing the method steps. Throughout the chapter reference is made to *Classic* STPA. This is the STPA method described within the *STPA Handbook* [96], and that forms the foundation of the *Shared Control* STPA method described below.

9.2 HIGHWAY ASSIST SYSTEM EXAMPLE

To help explain the hazard analysis method presented, a SAE Level 2 partial automation feature [141] has been chosen. This feature combines the longitudinal behaviour of ACC with lateral lane centring control. Being an SAE Level 2 feature means that both the vehicle's longitudinal and lateral control is shared between the human driver and the automation. But crucially, the human remains responsible for vehicle safety; potentially being called upon at a moments notice to maintain safety. Alfa Romeo's Highway Assist System (HAS) is such an example of a production system commercially available today [6, 25].

The HAS vehicle feature is used as an illustration here and will be revisited again later as part of the research evaluation.

The system uses on-board RADAR and cameras to detect the location and speed of preceding vehicles, and to determine the lane boundaries. The system also receives GPS and map data, which it uses to determine its current location and the nature of the road ahead. By determining its position on the map, the vehicle also knows whether it is within the ODD (i.e. a multi lane highway with a central reservation).

The use case for HAS is very similar to that of ACC. Once travelling on the highway, the Driver can decide to use HAS by pressing the appropriate steering wheel button while ACC is active. Having pressed the HAS enable button the system may take a number of seconds to initialise, but once initialised it will begin actively controlling the vehicle's lateral position in lane. The speedometer's circumference is illuminated in green to inform the Driver that HAS is now actively controlling the DDT. Providing the Driver maintains a light hand pressure on the steering wheel, the system will provide lateral and longitudinal vehicle control. The Driver can override the system at any time (e.g. to change lane), but may also receive requests from the system to intervene to maintain safety.

A use case description for HAS can be found in Appendix B. While the full HAS *Shared Control* STPA analysis can be found in Appendix C.

9.3 THE METHOD DESCRIBED

As a hazard analysis technique, STPA is becoming pervasive[1] within the automotive industry, particularly for AD systems. As introduced in Section 5.4, STPA requires the analyst to consider interactions between the system, human operators, and the environment. This makes STPA a more appealing choice

1 With 2022 seeing the introduction of the SOTIF Standard [165], which suggests STPA as an analysis technique, and the SAE STPA Recommended Practice [142] having been published, STPA's popularity is unlikely to wane within the automotive industry.

than methods such as FFA and FTA, given that the AD feature's safe behaviour is intrinsically linked to the correct perception and interpretation of the world in which it operates.

Other candidate analysis methods have been considered in relation to the EVCM, including FRAM [80] and EAST [185]. As discussed in Section 5.2.1.2, a benefit of the FRAM method is the ability to undertake meaningful analysis without the need for a system model. This is a powerful property, but it does make FRAM an incompatible method choice for use with the EVCM. Inspired by the ACC analysis by Banks and Stanton [19], this research also explored the EAST framework in relation to the EVCM. It was evident that the EVCM elements could help identify the EAST nodes, and the relationship between those nodes. However, how best to use the EAST framework and the EVCM together, to identify shared control hazard causes was not immediately apparent. It is for these reasons that STPA has been chosen as the basis of the EVCM's accompanying method.

Like *Classic* STPA (see Section 5.2.1.1), the *Shared Control* STPA method comprises four process steps (see Figure 12 on page 110):

Step ① define AD feature behaviour and agent responsibility conceptual model,

Step ② feature behaviour and shared control action modelling,

Step ③ unsafe control actions identification,

Step ④ shared control related loss scenario identification.

Step ① of this process is largely unchanged from *Classic* STPA, with the same approach being taken to define the purpose of the analysis. That is, identifying the system's stakeholders, potential losses and hazards. As an addition to *Classic* STPA, the behavioural competency taxonomy provides the language to describe the vehicle feature's behaviour.

Step ② involves the creation of a Control Structure Diagram (CSD). However, rather than developing this from first principles, a version of the EVCM is used as the CSD. Also, rather than analysing all potential control actions, the *Shared*

Control STPA method uses behavioural competencies to concentrate the analysis on those control actions whose successful control is the responsibility of more than one actor (i.e. shared between the human and the automation).

Having identified the control actions that are shared, Step ③ then identifies and determines which of those Unsafe Control Actions (UCAs) could be hazardous. This is achieved using the *Classic* STPA guide words of:

- *Not providing* a control action causes hazard.
- *Providing* a control action causes hazard.
- Providing a control action *too early, too late* or *out of sequence* causes hazard.
- Providing a control action *for too long* or *stopped too soon* causes hazard.

Step ④ then elicits loss scenarios that could lead to the UCAs and consequently the hazards identified earlier. Again Step ④ builds on the *Classic* STPA method, adding further questions that help the analyst elicit loss scenarios that may result because of human interactions [57] or the timing implications of those interactions [81].

9.3.1 *STEP 1: define AD feature behaviour and agent responsibility conceptual model*

Step 1 defines the purpose of the analysis, by identifying the stakeholders, losses and hazards. Not claiming to be an exhaustive list, the HAS stakeholders are identified as: the ego vehicle's driver and occupants, other road users, and the ego vehicle's manufacturer. From the stakeholder list the losses and hazards are identified, and these appear in Table 3. Although STPA refers to *losses* in its broadest sense (e.g. loss of security, loss of privacy), this research constrains its meaning to loss of life or financial loss. At this stage, the hazards and losses can be determined by a combination of engineering judgement and prior knowledge from similar systems.

Stakeholders
The Driver of the vehicle
Vehicle occupants
Other road users
Insurance companies
Vehicle manufacturer

Losses	
L1	Loss of life or injury to people
L2	Loss of or damage to vehicle
L3	Loss of or damage to objects outside the vehicle
L4	Loss of consumer confidence or sales
L5	Financial loss

Hazards		Potential Losses
H1	Vehicle does not maintain minimum safe distance to preceding / adjacent vehicles	L1, L2, L3, L4, L5
H2	Vehicle fails to stop for / avoid object in path	L1, L2, L3, L4, L5
H3	Succeeding vehicles cannot maintain minimum safe distance due to vehicle's deceleration rate	L1, L2, L3, L4, L5
H4	Vehicle manoeuvre leads to vehicle instability / loss of stability	L1, L2, L3, L4, L5

Table 3: Losses and hazards for a Highway Assist System (HAS)

Using the behavioural competency taxonomy[2] the behaviours implemented by the automation, together with those needed to maintain vehicle safety within the ODD, are identified. After which the responsibility for performing the applicable behavioural competencies is then determined (as described in Section 8.4 page 120). The behavioural competency responsibility table for HAS is shown in Table 4.

With HAS providing both lateral and longitudinal vehicle control, the manoeuvring level competencies immediately identifiable as being applicable are: *maintain position in lane*, *maintain speed*, and *follow other vehicles*. For HAS, the assumed behaviour for those three competencies is as described below, with all three being under the control of the automation while HAS is active:

- *Maintain speed:* HAS can sense the vehicle's current speed and understands the longitudinal acceleration rate (*Note:* negative rate implies a deceleration) needed to maintain the current desired vehicle speed.
- *Maintain position in lane:* HAS can sense the lane boundaries and the vehicle's relative position in lane. The system understands the vehicle's optimal position in lane (reference trajectory) needed to maintain a safe distance to static / dynamic objects, and decides what lateral adjustment is needed to follow the reference trajectory.

2 The complete behavioural competency taxonomy appears in Appendix A on page 210.

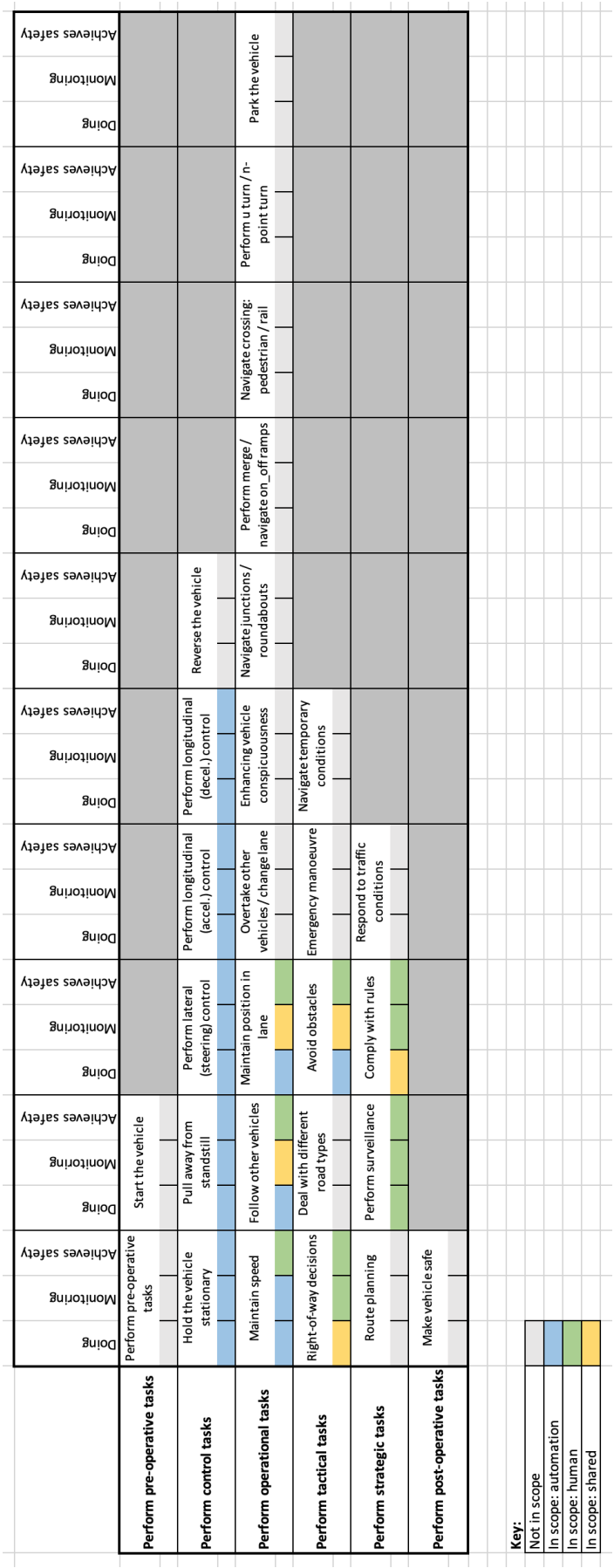


Table 4: Behavioural competencies applicable to HAS with actor responsibility identified

- *Follow other vehicles:* HAS can sense a preceding vehicle and understands the relative speed and distance to that vehicle. The system decides what changes are required to the vehicle longitudinal acceleration (*Note:* negative rate implies a deceleration rate) to maintain a consistent gap to the preceding vehicle.

If one then considers the hazards attributed to HAS (see Table 3 page 127), then the need for other behavioural competencies to mitigate the hazards identified becomes apparent. The behaviour competency *avoid obstacles* can sense object's in the vehicle path and understands which present a collision risk. The behavioural competency then decides what lateral adjustment and longitudinal acceleration rate changes are required to avoid a collision. Consequently, the *avoid obstacles* behavioural competency potentially mitigating hazard [H2] and perhaps also hazard [H1].

The behavioural competency *right-of-way decision* also potentially mitigates hazard [H2]. The *right-of-way decision* behavioural competency can sense traffic control devices, signage and infrastructure detail. It then uses this information to inform the correct vehicle trajectory and speed, which could include determining vehicle priority in relation to other road users. Being intended for use on multi-lane carriageways, the need for the *right-of-way decision* behavioural competency in relation to HAS may not be immediately apparent. However, several situations exist where the vehicle's current lane ends (e.g., a reduction in the number of lanes, or a lane closure on a Smart Motorway), which requires alterations to the vehicle's current trajectory and / or speed to maintain vehicle safety. Although the HAS automation might be capable of detecting some objects in path, the behavioural competencies *avoid obstacles* and *right-of-way decision* will be predominately the responsibility of the human driver.

Two further behavioural competencies are identified as having the potential to mitigate all hazards, by informing the correct use of automation. These are the *Strategic Level* behavioural competencies of *respond to traffic conditions* and *perform surveillance*, which are both the Driver's responsibility. The behavioural competency *respond to traffic conditions* can sense the vehicle's

operating environment (e.g., weather, visibility) and understands how these factors could influence the vehicle's performance and capability. For example, understanding that the HAS object detection performance might be reduced in foggy conditions. The *perform surveillance* behavioural competency can sense both the vehicle's environment, and the vehicle's behaviour in the environment. It understands how static and dynamic artefacts in the environment might affect vehicle safety, and decides what vehicle operation changes are needed to maintain safety. For example, the Driver might elect to slow down and stop using HAS when the traffic becomes heavy. Figure 15 shows the relationship between the competencies applicable to HAS.

There is an argument for including the *overtake / change lane* competency which potentially supports lateral control. For example, if *avoid obstacles* chooses to avoid the object in the road by changing lanes, rather than simply braking the vehicle to a stop. However, to avoid clutter, the potential to change lane is not modelled in Figure 15 on page 131. The full behavioural competency model, that does include the *overtake / change lane* competency together with all applicable control level competencies, can be found in Appendix C on page 251.

9.3.2 *STEP 2: feature behaviour and shared control action modelling*

Step 2 involves modelling the system's hierarchical control structure using a CSD. As introduced in Section 5.2.1.1, Leveson and Thomas suggest that abstraction is used to manage complexity. Using the EVCM as the basis for the CSD, maintains a sufficiently high level of abstraction for the initial analysis.

The CSD drawing convention is to place the controller with the highest control authority at the top. Consequently, a CSD for an automated driving system will typically have a controller box at the top labelled "the driver". However, modelling the driver in this way would not reflect the shared nature of HAS control. In contrast, the EVCM does facilitate the modelling of the shared control

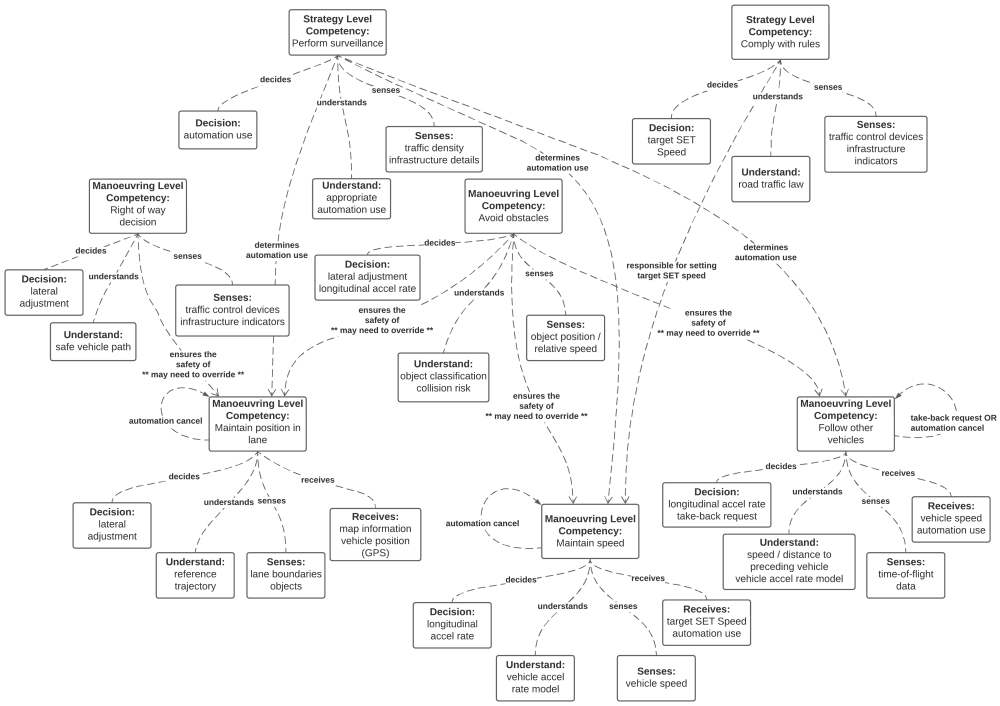


Figure 15: Behavioural competencies applicable to HAS and their relationship to one another

taking place between the driver and the automation. Therefore, by redrawing the EVCM in the *'STPA style'*[3] the shared nature of control can be explored.

With the EVCM providing the CSD for the analysis, the next task is to use the behavioural competencies identified in Step 1 to determine the control actions and applicable information feedback paths for HAS. Capturing the relationships between behavioural competencies (as one might do in an ontology diagram), Figure 15 identifies the two control actions influenced by the HAS automation (i.e. *lateral adjustment* and *longitudinal accel rate*) and illustrates how the decision making regarding those control actions is due in part to other competencies in the control hierarchy. Namely, *avoid obstacles*, *right-of-way decision*, *comply with rules* and *perform surveillance* are, with the exception of *avoid obstacles*, the sole responsibility of the human driver. The complete CSD for HAS appears in Figure 16.

3 That is, the controlling processes hierarchically placed above the controlled process, and the elements representing perception, understanding and error incorporated at each hierarchical level.

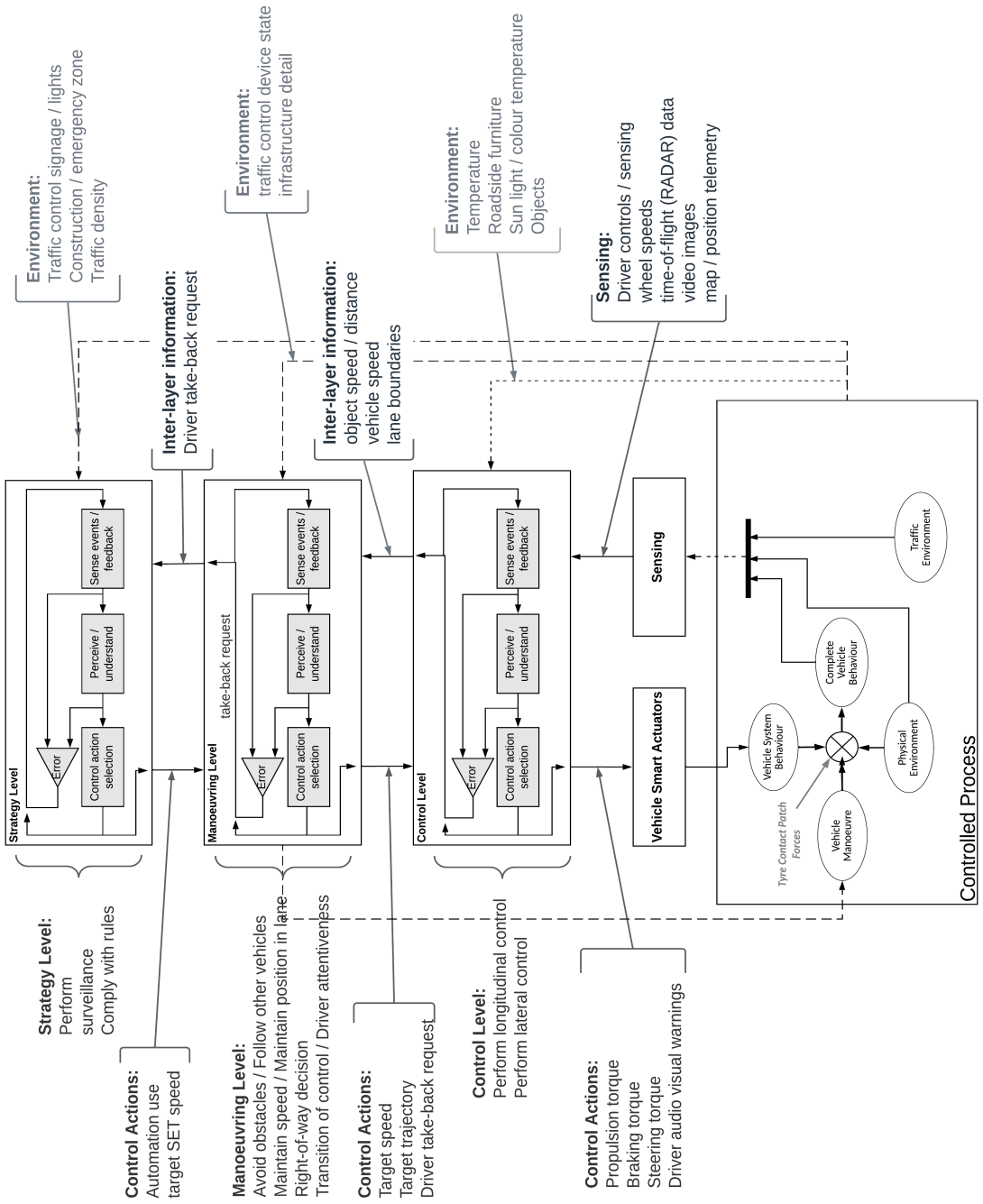


Figure 16: Populated CSD for HAS

Having identified *lateral adjustment* and *longitudinal accel rate* as having shared control, these are taken forward into Step 3.

9.3.3 *STEP 3: Unsafe control action identification*

Step 3 identifies the Unsafe Control Actions (UCAs) for the system. A control action is deemed unsafe when, if not constrained by appropriate safety measures, it would lead to unsafe system behaviour and ultimately losses. *Classic* STPA applies the guide words (see Section 9.3 page 124) to each control action in turn. The same is done here, although each control action is also considered in the context of each behavioural competency. For example, *avoid obstacles* does not provide sufficient *lateral adjustment* to avoid a collision with the object in the vehicle's path (Hazard H2). Or the *follow other vehicles* behaviour provides a large *deceleration rate*, which affords the vehicle behind insufficient time to avoid a collision (Hazard H3). Considering potential interactions between competencies also uncovers further UCAs. For example, *maintain position in lane* provides *lateral adjustment (centring force)* while *right-of-way decision* is also applying a lateral adjustment to change lane to avoid a collision with a vehicle in path (Hazard H2). Using this technique the UCAs for the HAS CSD are identified, and these appear in the table on page 134.

9.3.4 *STEP 4: Loss scenario identification*

Step 4 identifies loss scenarios for the UCAs identified in Step 3. Having identified the loss scenarios, then system safety requirements can be specified to mitigate each loss scenario. Step 4 adds further loss scenario types to *Classic* STPA (see Section 5.2.1.1 page 76). These further loss scenario types have been added to address situations where either a control action is influenced by multiple competencies, or the responsibility for a behavioural competency is transferred.

Control Action	Not providing causes hazard	Providing causes hazard	Too early, too late, out of order causes hazard	Stopped too soon, applied too long causes hazard
Lateral adjustment CONTROLLED BY: *Maintain position in lane* *Avoid obstacles* *Right-of-way decision*	**UCA-1:** *maintain position in lane* does not provide *lateral adjustment* to maintain central position in lane [H1]	**UCA-4:** *maintain position in lane* provides *lateral adjustment* when no path deviation is required [H1]	"Too early" considered the same as "providing"	"Stopped too soon" considered the same as "not providing"
	UCA-2: *avoid obstacles* does not provide *lateral adjustment override* when object present in path [H2]	**UCA-5:** *maintain position in lane* provides *lateral adjustment (centring)* while *avoid obstacles* is also applying *lateral adjustment* to avoid an object in path [H2]	"Too late" considered the same as "not providing"	"Applied too long" considered the same as "providing"
	UCA-3: *right-of-way decision* does not provide *lateral adjustment override* when path deviation required (i.e. loss of lane) [H2]	**UCA-6:** *maintain position in lane* provides *lateral adjustment (centring)* while *right-of-way decision* is also applying a *lateral adjustment* to change lane [H2]	"Out of order" considered the same as "providing"	
		UCA-7: *avoid obstacles* provides *lateral adjustment* while *maintain position in lane* is also applying a *lateral adjustment* [H1, H4]		
		UCA-8: *right-of-way decision* provides *lateral adjustment* while *maintain position in lane* is also applying a *lateral adjustment* [H4]		
(Negative) Longitudinal acceleration rate (i.e. deceleration) CONTROLLED BY: *Maintain speed* *Follow other vehicles* *Avoid obstacles*	"Not providing" not applicable to *maintain speed.*	**UCA-11:** *maintain speed* provides a large *deceleration rate* change with vehicle close behind [H3]	"Too early" considered the same as "providing"	"Stopped too soon" not considered applicable to *maintain speed*
	UCA-9: *follow other vehicles* does not provide *deceleration rate* change when preceding vehicle slows [H1]	**UCA-12:** *maintain speed* provides a large *deceleration rate* change on low mu surface [H4]	**UCA-17:** *avoid obstacles* provides negative *acceleration (deceleration) rate* change too early or late with object in vehicle path [H1, H2]	"Stopped too soon" considered the same as "not providing" for *follow other vehicles* and *avoid obstacles*
	UCA-10: *avoid obstacles* does not provide *deceleration rate* change with object in vehicle path [H1, H2]	**UCA-13:** *follow other vehicles* provides a large *deceleration rate* when vehicle close behind [H3]	**UCA-18:** *avoid obstacles* provides negative *acceleration (deceleration) rate* change too late following a Driver take-back control request being issued by *follow other vehicles* [H1, H2]	"Applied too long" considered the same as "providing"
		UCA-14: *follow other vehicles* provides a large *deceleration rate* change on a low mu surface [H4]	"Too late" not considered applicable to *maintain speed*	
		UCA-15: *avoid obstacles* provides a large *deceleration rate* with vehicle close behind [H3]	"Too late" considered same as "not providing" for *follow other vehicles*	
		UCA-16: *avoid obstacles* provides a large *deceleration rate* on low mu surface [H4]	"Out of order" not considered applicable	

Table 5: UCAs for the HAS vehicle feature

The behavioural competency interaction diagram for HAS (Figure 15 on page 131) illustrates how the behavioural competencies of *maintain position in lane*, *avoid obstacles*, and *right-of-way decision*, might simultaneously influence the *lateral adjustment* control action. An example of this might be where HAS is actively controlling the vehicle's position in lane and the Driver applies a steering input to override the system, perhaps because the Driver has identified an object in the vehicle's path (i.e., the *avoid objects* behavioural competency), or because the lane ahead has been closed off with traffic cones (i.e., the *right-of-way decision* behavioural competency).

For partial or conditional automation there will be situations where responsibility for a behavioural competency passes from human to the automation and vice versa. An example for HAS might be the responsibility for the behavioural competency *maintain position in lane* transferring between the automation and the human driver during an overtaking manoeuvre. Differences in perception between the human and the automation, and the time available to complete the transition (see the *cyclical model of human action* discussion in Section 5.2.2.1 on page 80), both potentially affect a successful transition of control.

To help the analyst consider the loss scenarios relevant to a given UCA, a set of guiding questions has been developed (see Figure 17 page 137). This list of loss scenario questions has been generated by reviewing and correlating the factors identified by Leveson and Thomas in *STPA Handbook* as developing loss scenarios [96], with the factors identified by France as affecting the human operator's mental models[4] [57]. In addition to the factors that result from incorrect, incomplete or out-of-date mental models, the temporal aspect of the HMI proposed by Hollnagel and Wood, is also considered [81]. This is particularly relevant when the human is expected to make a correct and timely response to a system warning or instruction. Therefore, questions relating to the cyclical model of human action (see Section 5.2.2 page 79) are added to

4 In addition to the *Classic* STPA loss scenario types, France suggests that loss scenarios may result from incorrect control action selection (by the human operator), and from errors in the human operators mental models. With the additional loss scenario guide questions covering misunderstandings regarding the process state, the process behaviour and the operating environment, and the process by which those mental models are updated [57].

prompt the identification of loss scenarios associated with the temporal nature of shared control.

9.3.4.1 *Lateral adjustment related loss scenarios*

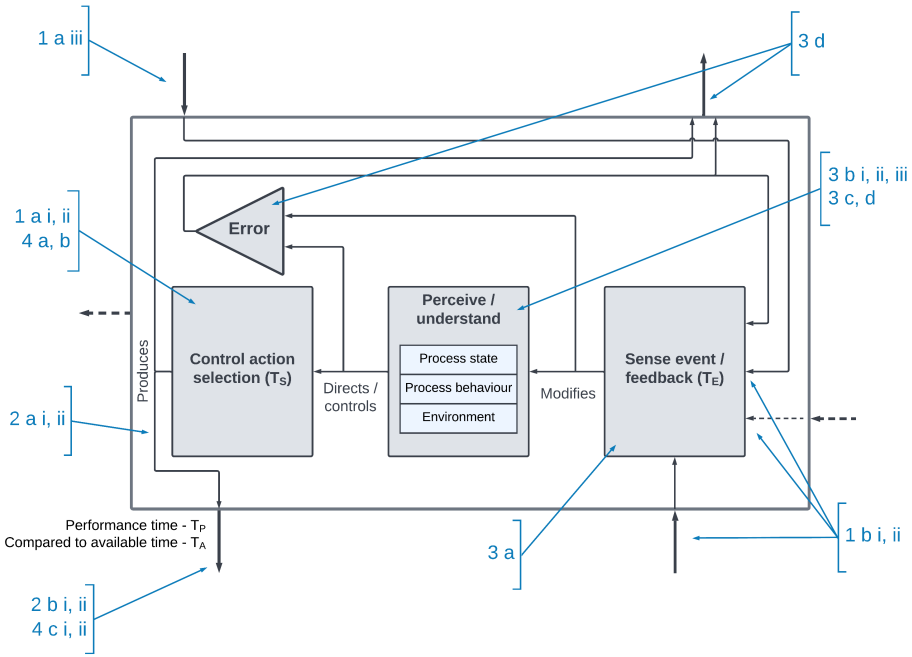
Considering correct *lateral adjustment* first, then the loss scenario questions from Figure 17 are asked in relation to the Unsafe Control Actions: UCA–1 to UCA–8. To demonstrate this process step, the loss scenario questions have been applied for UCA–1 and UCA–7 below:

- UCA–1: *maintain position in lane* does not provide *lateral adjustment* to maintain safe position in lane [H1]
- UCA–7: *avoid obstacles* provides *lateral adjustment* while *maintain position in lane* is also applying a *lateral adjustment* [H1, H4]

The application of the loss scenario questions to UCA–1 and UCA–7 are shown pictorially in Figure 18 (a) and (b) respectively.

Classic STPA would ask loss scenario questions about the controller and the information / feedback it receives (Questions 1 Figure 17) and the plant behaviour in response to control actions (Questions 2 Figure 17). Applying such loss scenario questions to UCA–1 uncovers scenarios relating to the controller, its inputs and control of the vehicle itself. For example, failures within the control algorithm or vision system which result in either the *lateral adjustment* being applied incorrectly or the lane boundary being detected incorrectly. Both of which could potentially result in the vehicle departing its lane and either hitting road-side furniture or a vehicle in an adjacent lane. Examples relating to vehicle control include: no consideration being given to the road adhesion available resulting in the required lateral adjustment not being achieved, and out of date map data being received which means the system does not inhibit HAS use while the vehicle is travelling through a road works zone. As with the above examples, the likely consequence is again the vehicle straying out of its lane and either hitting road-side furniture or another road user.

Asking the loss scenario questions relating to perception and understanding (Question 3 Figure 17) uncovers further loss scenarios. For example, having



'Classic' STPA

1. Identifying scenarios that lead to unsafe control actions
   a. unsafe controller behaviours
      i. failures involving the controller
      ii. inadequate control algorithm
      iii. unsafe control input (from another controller)
      iv. inadequate process model
   b. causes of inadequate feedback and information
      i. feedback or information not received
      ii. inadequate feedback is received

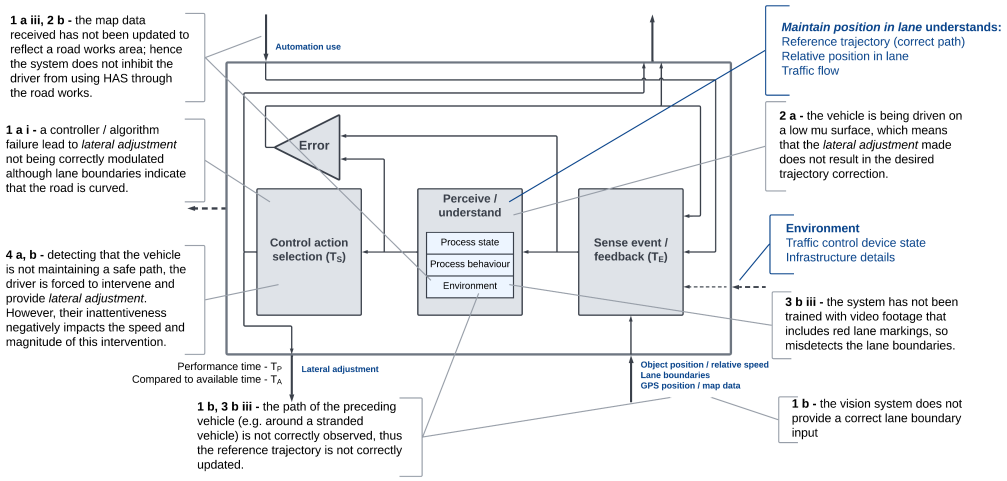
2. Identifying scenarios in which control actions are improperly executed or not executed
   a. scenarios involving the control path
      i. control actions not executed
      ii. control actions improperly executed
   b. scenarios related to the controlled process
      i. controlled process does not respond to the control action
      ii. controlled process responds as though a control action has been applied (when it has not)

'Shared Control' STPA

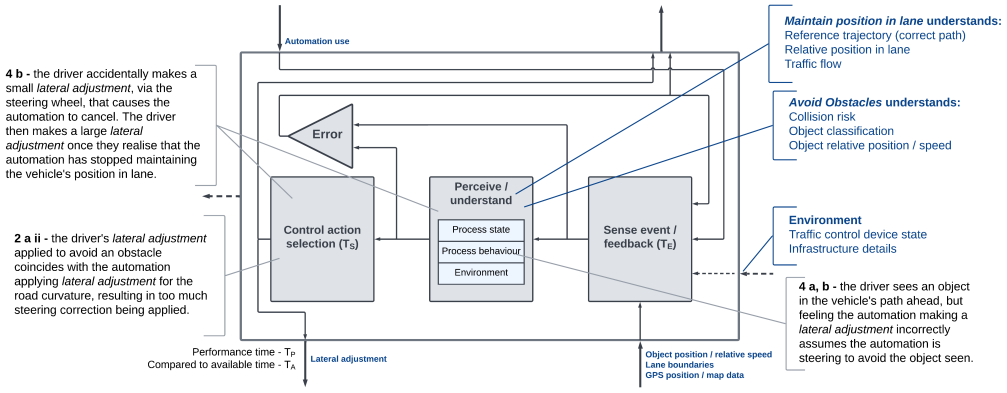
3. Event evaluation and construct maintenance
   a. What could impact an agents reactions / time needed to evaluate an event (T_E)
   b. What does the agent know or believe about the system
      i. process state
      ii. process behaviour
      ii. environment
   c. How did the agent come to have their current knowledge or beliefs (mental model updates)
   d. How might a difference in expected behaviour (error) influence what the agent does next

4. Choice and execution of control action
   a. How did the agent choose which control action to perform
   b. What could impact an agents proactive actions / timely selection of control actions (T_S)
   c. In relation to the controlled process how much time is
      i. needed to complete the control action (T_P)
      ii. available to perform the control action (T_A) in the context of the controlled process

Figure 17: Focused loss scenario identification questions (originating from [96] [57] [81])



(a) UCA–1



(b) UCA–7

Figure 18: Potential loss scenarios for unsafe control actions UCA–1 and UCA–7

used video footage containing examples of yellow and white road marking to train the vision system, might mean that the system is unable to correctly detect lane boundaries in locations where red road markings are used. Again such a loss scenario would result in the vehicle not maintaining the correct position in lane. This then leads into the control action selection loss scenario questions (Question 4 Figure 17). For example, detecting that the vehicle has begun to stray from its lane, while HAS is operational, the driver might intervene and begin steering. However, because the Driver may have become somewhat inattentive the speed and magnitude of their steering correction is too large, which causes the vehicle to become unstable. This again uncovers a loss scenario where the vehicle exits its lane and potentially hits something or someone.

Being focused on the potential interaction between behavioural competencies (*avoid obstacles* and *maintain position in lane*) the potential loss scenarios for UCA–7 are more driver centric than UCA–1. Considering the plant's behaviour in response to a control action (Question 2 Figure 17) highlights a loss scenario where both the Driver and the automation attempt to apply a *lateral adjustment* simultaneously. The consequence of which is a lateral vehicle response that is too large, leading to either a departure from lane or to vehicle instability.

Similarly, asking loss scenario questions relating to correct and timely control action selection (Question 4 Figure 17) uncovers two further loss scenarios relating to interactions between the driver and the automation: In the first example, the Driver has seen an object in the road far in the distance. When the Driver then detects HAS making a *lateral adjustment*, felt as feedback through the steering wheel, they assume that the automation has detected and hence is steering to avoid the obstacle (i.e. an incorrect mental model of the process behaviour). Or, while HAS is active the Driver makes an unintentional steering input, which the automation interprets as the Driver wishing to override the system. This intervention by the Driver results in the HAS feature cancelling and relinquishing its DDT control. In either case, the likely outcome is the vehicle departing from its lane and potentially hitting road-side furniture or another road user.

Having identified the loss scenarios in which the *lateral adjustment* UCAs lead to hazards, safety requirements can then be written for each of the loss scenarios identified. For example, having identified that a small accidental *lateral adjustment* by the Driver could result in HAS cancelling unintentionally, a safety requirement could be written to specify the magnitude and duration of the Driver intervention needed to cancel HAS. A second safety requirement specifying that the HAS current state (i.e. controlling vehicle lateral position or not) be obvious to the Driver at all times, would make HAS's operating state explicit.

9.3.4.2 *Longitudinal acceleration rate related loss scenarios*

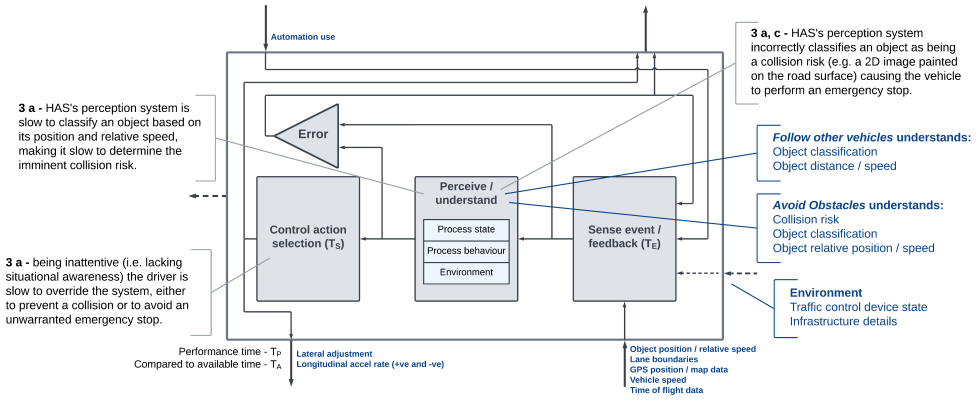
The same process can then be applied to the UCAs associated with *longitudinal acceleration rate* (i.e. UCA–9 to UCA–18), or more specifically a negative acceleration when the automation is expected to slow the vehicle. The process is applied to UCA–17 and UCA–18 here to demonstrate the temporal aspect of the enhanced loss scenario questions:

- UCA–17: *avoid obstacles* provides negative *longitudinal acceleration rate* change (deceleration) too early or late with object in vehicle path [H1, H2]
- UCA–18: *avoid obstacles* provides negative *longitudinal acceleration rate* change (deceleration) too late following a Driver take-back control request being issued [H1, H2]

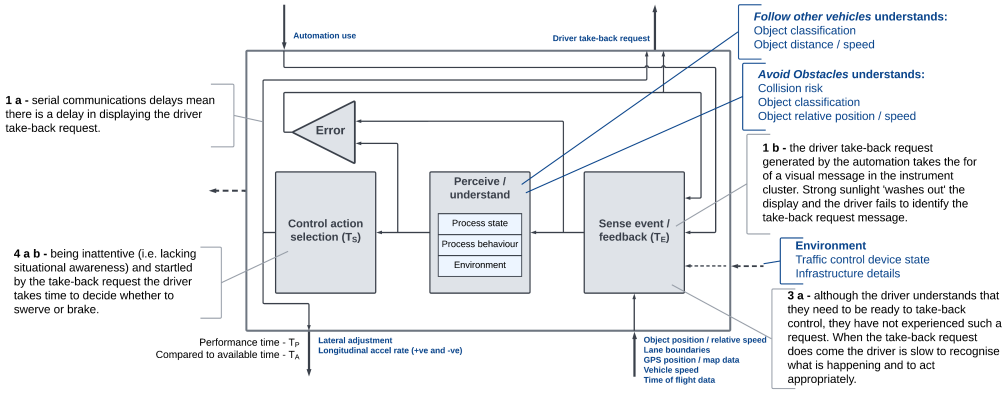
The application of the loss scenario questions to UCA–17 and UCA–18 are shown pictorially in Figure 19 (a) and (b) respectively.

Figure 19 (a) highlights the loss scenarios for UCA–17, by exploring the loss scenarios in which the *avoid obstacles* behavioural competency either achieves the intended objective too early or too late. Although responsibility for the *avoid obstacles* behavioural competency is potentially shared, UCA–17 considers loss scenarios where the automation responsible for undertaking the competency. In the context of HAS controlling longitudinal acceleration too early or too late is interpreted as the system either failing to avert a collision with an object in path that poses a collision risk, or the system bringing the vehicle to an emergency stop when no collision risk exists. This second case puts the vehicle at increased risk of being hit by the following vehicle. The automation control algorithm misclassifying the type or position of an object are loss scenario examples where the automation's control of the *longitudinal acceleration rate* control action is potentially unsafe. While HAS is active, this error in system perception is potentially compounded by the Driver losing situational awareness and being slow to override the automation to maintain vehicle safety.

Perhaps the key difference for *longitudinal acceleration rate* control is the potential for the automation to request that the Driver take-back control. This



(a) UCA–17



(b) UCA–18

Figure 19: Potential loss scenarios for unsafe control actions UCA–17 and UCA–18

might occur while HAS is undertaking the *follow other vehicles* behavioural competency, if the automation detects that the headway to the preceding vehicle has diminished rapidly, and it requires the Driver to make an emergency brake application. This might occur because a vehicle from an adjacent lane cuts-in in front of the vehicle, or because the vehicle ahead cuts-out to reveal a stationary vehicle ahead. In Figure 19 (b) this Driver take-back request is modelled as an internal feedback loop within the *Manoeuvring Level* block.

Figure 19 (b) shows the loss scenario questions relating to event reaction time (see Question 3 Figure 17) and control action selection (see Question 4 Figure 17) being applied for UCA–18 in the context of the take-back request. In the first example, the Driver understands that they might be required to take-back DDT control from the automation, but having never experienced such a request, when the request does come they are somewhat startled and slow to react.

Similarly, having been driving with HAS active for sometime, the Driver has become inattentive and lacks situational awareness. Consequently when the automation does issue a take-back request the Driver takes time to determine whether changing lane or slowing down is the most appropriate course of action to take.

When considering take-back requests in the context of the *avoid obstacles* behavioural competency one immediately thinks about the response of the Driver, and hence Questions 3 and 4. However, asking Questions 1 and 2 (see Figure 17) may also uncover potential loss scenarios. For example, there may be loss scenarios where the automation's behaviour could causes the take-back request not to occur, or to occur late (i.e. Question 1 Figure 17). Or there could be errors with the HMI which means that the take-back request is not seen by the Driver. Perhaps the take-back request is a visual warning in the instrument cluster, which is washed-out by strong sunlight.

9.4 SUMMARY

The *Shared Control* STPA method presented follows the same general format as *Classic* STPA, but includes a number of important differences. Firstly, rather than creating the CSD for the candidate AD feature from scratch, the CSD is created from the hierarchical controller style EVCM. This ensures the analysis is focused at an appropriate level of abstraction from which to explore shared control hazard causes.

The behavioural competency taxonomy then provides the vocabulary to describe the AD feature's behaviour in the context of its intended ODD. The importance of this to the safety analysis is twofold: Firstly, those behavioural competencies needed to maintain vehicle safety within the ODD can be identified explicitly. Secondly, once identified, the agent having responsibility for undertaking (the *doing*), *monitoring* and maintaining vehicle safety (*achieves safety*) can then be captured. Once complete, the behavioural competency taxonomy matrix provides a rich definition of the candidate AD feature's shared

control, and helps focus the subsequent analysis towards those control actions directly influenced by shared control.

Applying the STPA guide words (see Section 9.3 page 124) in the context of combinations of behavioural competencies adds another dimension to the UCA analysis, by helping the analyst to identify UCAs resulting from potential shared control conflicts. For example, considering the behavioural competency *maintain position in lane* together with *avoid obstacles* uncovers the UCAs that might exist because of such a conflict – i.e., while the machine agent is endeavouring to keep the vehicle centred in lane, the driver is desperately trying to steer the vehicle around the large object that has just fallen off the back of the preceding vehicle.

Finally, creating the loss scenario question list (see Figure 17 page 137) consolidates the loss scenario analysis. By considering each question in turn, for all identified UCAs, the analyst can be confident that the spectrum of loss scenario types have been explored, namely: losses resulting from failures in control, losses resulting from an incorrect response by the controlled plant, losses due to failures in the formation and maintenance of mental models, or failures due to the temporal nature of shared control.

Together these steps provide an expansive examination of shared control for the candidate AD feature. As will be shown in Chapter 10, when used to populate a safety case argument pattern, the resulting evidence builds a shared control confidence argument for the AD feature's safety case.

10

SHARED CONTROL SAFETY CASE ARGUMENT

10.1 INTRODUCTION

This chapter follows a similar process structure to that used in the Assurance of Machine Learning for use in Autonomous Systems (AMLAS)[1] methodology [11]. By providing a process and set of argument patterns that when instantiated as part of the AD vehicle feature's development, this chapter presents a safety case argument pattern for shared control. When used as part of a broader AD feature safety case, this safety case argument method helps to produce a reasoned argument for the shared control aspects of the AD feature.

As mentioned in Section 7.3 both ISO 26262 [162] and ISO 21448 [165] require the creation of a safety case argument for a vehicle feature development. Neither standard prescribes a particular safety case argument format, but the Goal Structuring Notation (GSN) is the notation frequently used by engineering argument practitioners [161]. It is for this reason that the shared control safety case argument pattern presented here also uses GSN. GSN is a graphical notation that allows claims, evidence and context relating to a given safety case argument to be explicitly conveyed. Significantly, the relationship between the safety case argument elements can also be conveyed explicitly using GSN [161]. Signified by braces (i.e., { }), the GSN patterns presented within this chapter are generic uninstantiated *patterns*. The expectation is that the automotive safety practitioner will take these argument patterns for shared control and instantiate the terms in braces with their own vehicle feature.

1 AMLAS integrates safety assurance into the development of machine learning components, through the provision of a process and a set of safety case patterns. The expectation is that once instantiated, the argument patterns will form part of a complete system safety case [11].

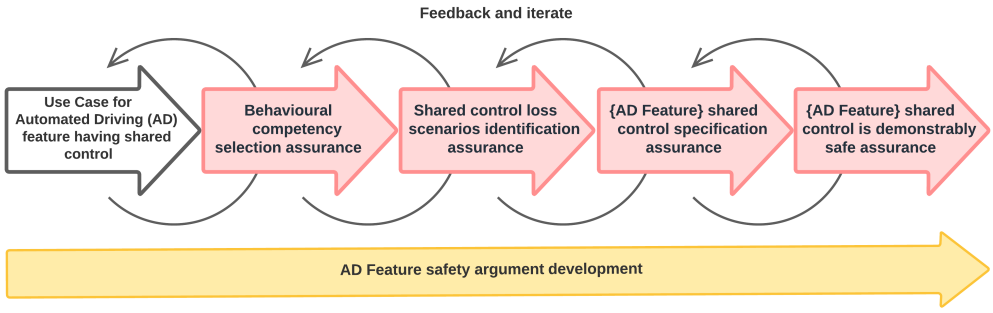


Figure 20: An overview of the shared control safety case argument methodology

This chapter first describes each step that comprises the safety case argument methodology. The chapter concludes by discussing how one might incorporate the elements of the shared control safety case argument into the broader vehicle feature development. This includes incorporating both the instantiated safety case argument pattern into the wider safety case argument, but also incorporating artefacts, such as shared control related safety requirements and test specifications, into the broader engineering development process.

10.2 SHARED CONTROL SAFETY CASE ARGUMENT METHODOLOGY

10.2.1 *Overview*

An explicit safety case argument for the shared control aspects of an AD vehicle feature is developed by following the 4 steps shown in Figure 20 and described in the sections below. The process starts with a feature definition. Typically this will be a use case description[2] of the feature functionality and will describe how the user of the system (in our case, the vehicle driver) interacts with the feature. The expectation is that the safety case development will happen in parallel with the AD feature development. Although shown as a single process flow here, in reality the process is iterative; with discoveries made during a given step having the potential to impact prior steps.

2 Similar to those use case examples given in Appendix B.

The below sections describe the objectives of each assurance step, together with the inputs and outputs to that step. A description of the assurance step is also included, with the activities that might be undertaken and decisions made during that step discussed. Where applicable the argument pattern is also included.

10.2.2 *Top level safety case argument*

Intended to form part of a larger AD feature safety case argument, Figure 21 shows the top goal for the shared control safety case argument. The main claim (G1) is that vehicle accidents caused by the shared control between the human driver and the automation, while the {AD Feature} is being used, are avoided. This goal is supported by three further claims: firstly, that the *Shared Control* STPA method identifies all UCAs and loss scenarios that are the result of the {AD Feature's} shared control (claim G2), secondly, that safety requirements have been written to mitigate those UCAs and loss scenarios identified (claim G3), and thirdly that loss scenario informed scenario based testing demonstrates the safe behaviour of the {AD Feature's} shared control (claim G4). Goals G2, G3 and G4 are further developed in the following steps.

10.2.3 *Assurance step 1: behavioural competency selection assurance*

10.2.3.1 *Objectives*

Assurance Step 1 (shown pictorially in Figure 22) comprises four activities and seeks to achieve the following objectives:

1. To define the behavioural competencies undertaken by the automation
2. To define the additional behavioural competencies needed to maintain vehicle safety while the automation is active and
3. To instantiate the EVCM, having defined the behavioural competencies and shared control actions.

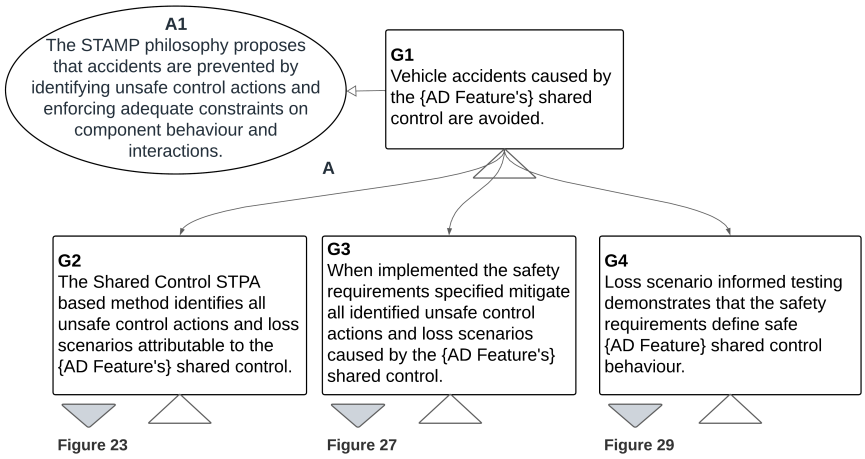


Figure 21: Goal 1: an explicit safety case argument pattern for the shared control aspects of an AD vehicle feature.

10.2.3.2 *Inputs and outputs*

Assurance Step 1 requires two inputs: a use case describing the {AD Feature} and the behavioural competency taxonomy (see Appendix A). A detailed description of the feature behaviour under all operating conditions is not required to begin this step. However, the expected behaviour during normal operation (e.g. the use case's *Main Flow*), expected alternate scenarios (e.g. different *Alternate Flows* for the various driving scenarios or operating modes expected) and failures (e.g. *Exception Flows* for component failure or loss of critical data) are needed. The second input is the behavioural competency taxonomy. The taxonomy includes all identified competencies necessary to undertake the DDT. This includes both competencies associated with vehicle control (e.g. *maintain position in lane*), but also those needed to maintain safe vehicle operation (e.g. *avoid obstacles*).

This step generates two output artefacts: a behavioural competency interaction diagram and the instantiated EVCM; with both outputs being key inputs of the hazard analysis process.

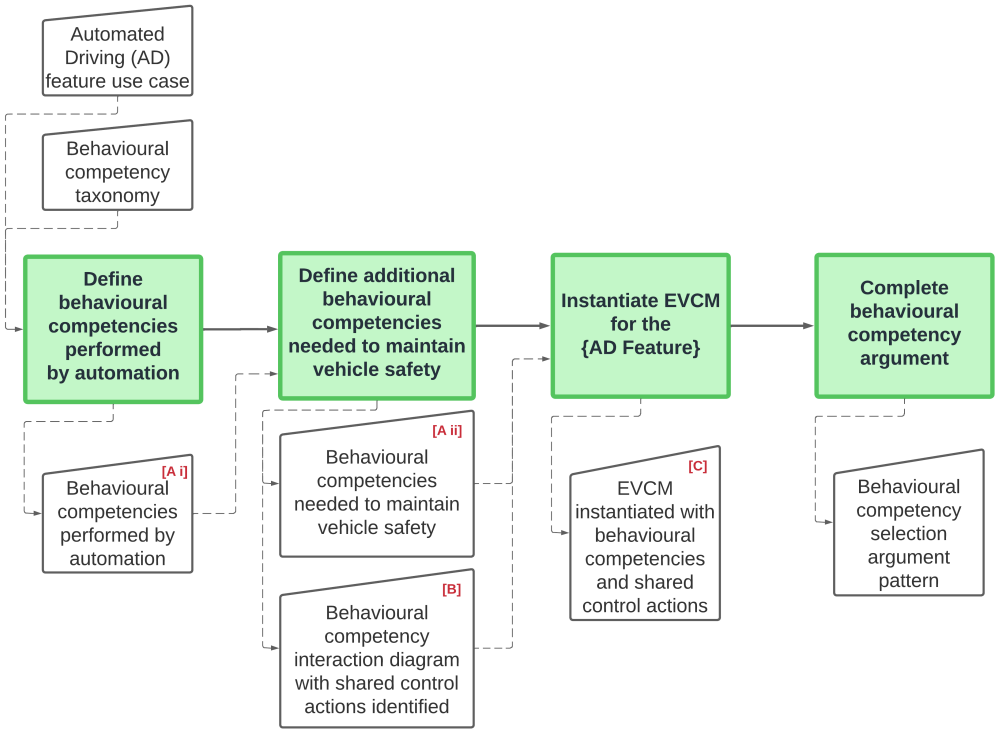


Figure 22: Assurance Step 1: Behavioural competency selection assurance process flow

10.2.3.3 *Activities*

Assurance Step 1 comprises the following activities:

Activity 1: Behavioural competencies performed by the automation.

The first input to this activity is the {AD Feature} use case. The use case describes the {AD Feature}'s functional operation together with the expected interaction between the automation and the human driver. The second input is the behavioural competency taxonomy. The activity considers the {AD Feature} functionality in the context of the behavioural competency taxonomy to identify the competencies that will be undertaken by the automation while it is operational (Figure 22 artefact [A i]).

Activity 2: Define additional behavioural competencies needed to maintain vehicle safety. These are the competencies, such as *avoid obstacles* and *right of way decision*, that maintain vehicle safety while the automation is operational (Figure 22 artefact [A ii]). These competencies are combined

with the competencies identified above (Figure 22 artefact [A i]), to form a behavioural competency interaction diagram for the {AD Feature}. The behavioural competency interaction diagram is the first output from Assurance Step 1, and is a pictorial representation of the interaction between competencies delivered by the {AD Feature} functionality, and the additional competencies needed to maintain vehicle safety (Figure 22 artefact [B]).

Activity 3: Instantiate EVCM for the {AD Feature}.

Pictorially representing the relationship between competencies delivered by the automation (identified during Activity 1) and those needed to maintain vehicle safety (identified during Activity 2) helps uncover the shared control actions. Activity 3 takes these shared control actions, together with the identified behaviour competencies and instantiates the EVCM. The instantiated EVCM is the second output from Assurance Step 1 and is a key input to the hazard analysis process (Figure 22 artefact [C]).

Activity 4: Complete behavioural competency argument.

The Assurance Step 1 activities support the loss scenario identification claim (see G2 in Figure 23) by creating the material necessary to complete the downstream analysis. Identifying the behavioural competencies undertaken by the automation, together with those necessary to maintain vehicle safety while the automation is in use (evidence [A i] and [A ii] respectively), and representing those competency interactions pictorially (evidence [B]), uncovers the nature of the {AD feature}'s shared control – claim G2.1. Instantiating the EVCM with the identified behavioural competencies and shared control actions (evidence [C]) forms the prerequisite material needed to complete the *Shared Control* STPA hazard analysis method (claim G2.3), which is the focus of Assurance Step 2.

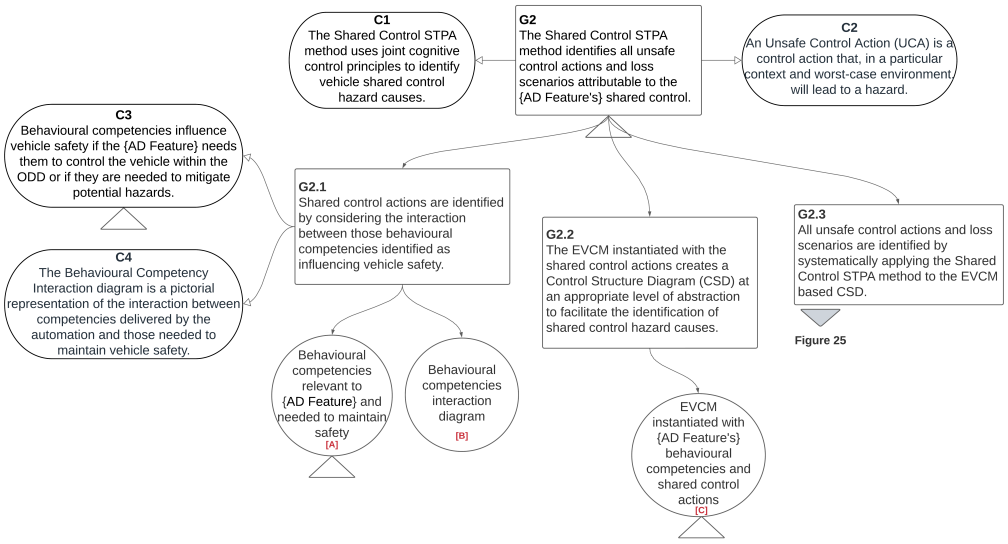


Figure 23: Goal 2: shared control unsafe control action and loss scenario identification assurance argument pattern

10.2.4 *Assurance Step 2: shared control loss scenarios identification assurance*

10.2.4.1 *Objectives*

Assurance Step 2 (shown pictorially in Figure 24) comprises four activities and seeks to identify all potential loss scenarios attributable to shared control, through the systematic application of the *Shared Control* STPA method. In doing so, this step seeks to fulfil the following objectives:

1. To identify the stakeholders, losses and hazards applicable to the {AD Feature}
2. To identify the UCAs pertaining to shared control through the systematic application of STPA guide words.
3. To identify those loss scenarios resulting from errors in control, feedback and control execution (*Classic* STPA) and
4. To identify those loss scenarios resulting from errors in event evaluation, understanding and response time (*Shared Control* STPA).

10.2.4.2 *Inputs and outputs*

Assurance Step 2 has a number of inputs. The {AD Feature} use case definition is again used here, this time to identify the relevant stakeholders, the stakeholders' losses and likely system level hazards. The second input are the STPA guide words (see Section 9.3 page 124) used to systematically identify the UCAs from the shared control actions identified during Assurance Step 1. The behavioural competencies interaction diagram (created during Assurance Step 1) also facilitates the identification of UCAs. The final set of inputs are the instantiated EVCM from Step 1 and the question list (see Figure 17 in Section 9.3.4) which supports the discovery of *Classic* and *Shared Control* loss scenarios types.

Assurance Step 2 outputs the following lists: the stakeholders relevant to the {AD Feature}, the losses (e.g. accidents) that those stakeholders might experience, the potential hazards caused by the {AD Feature}, the systematically identified UCAs associated with shared control, and the potential loss scenarios attributable to shared control.

10.2.4.3 *Activities*

Assurance Step 2 comprises the following activities:

Activity 1: Identify the stakeholders, losses (accidents) and hazards.

The {AD Feature} use case helps identify the relevant stakeholders (Figure 24 artefact [D]), the losses those stakeholders might experience (artefact [E]), together with the probable hazards (artefact [F]). In an automotive context the stakeholders will undoubtably include the vehicle's occupants (i.e. the driver and passengers), and other road users, but could also include the vehicle's manufacturer and insurance underwriter. The losses specified will be associated with the identified list of stakeholders and will typically be phrased in terms of *harm*, *damage* or *financial loss*. STPA defines a hazard as *"a system state or set of conditions that, together with a particular set of worst-case environmental conditions, will lead to a loss"* [96]. In an automotive context, hazards will typically involve describing system states that could lead to a loss of vehicle control (e.g. undemanded

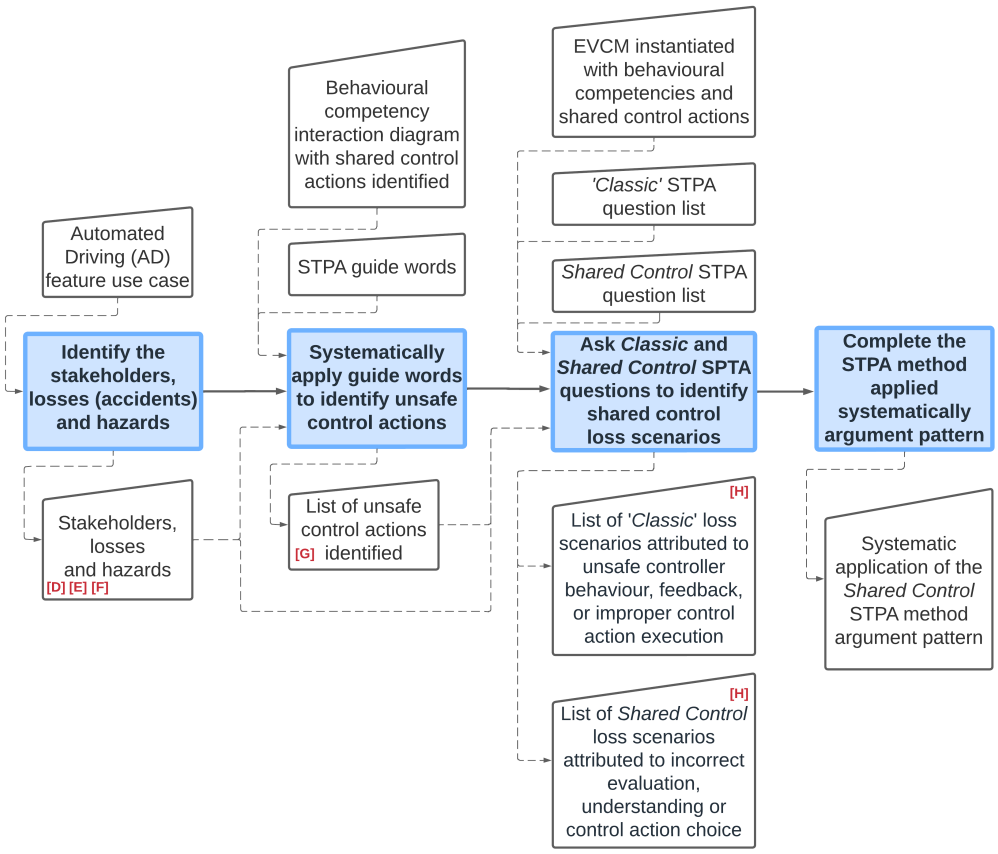


Figure 24: Assurance Step 2: the focused analysis of shared control assurance process flow

acceleration) or conditions that in themselves are hazardous (e.g. vehicle does not maintain a safe distance to the preceding vehicle).

Activity 2: Systematically apply guide words to identify UCAs.
This activity uses the STPA guide words and the behavioural competency interaction diagram (artefact [B] from Assurance Step 1) as its input. Using the STPA guide words, and informed by the behavioural competencies identified, all potential UCAs are discovered (Figure 24 artefact [G]).

Activity 3: Ask *Classic* and *Shared Control* STPA questions to identify shared control loss scenarios. Using the instantiated EVCM and the list of UCAs, this activity uses guided questioning to uncover loss scenarios (Figure 24 artefact [H]) falling into one of four loss scenario types: control type loss scenarios that result from errors in control or feedback, execution type loss scenarios that result from errors in the execution of correctly applied control actions, evaluation type loss scenarios that result when the controller incorrectly evaluates an event, or its understanding is incorrect, and incorrect action choice loss scenarios where an incorrect control action is chosen.

Activity 4: Instantiate the systematic application of a Shared Control STPA process argument pattern. Figure 25 contains the safety case argument pattern for Assurance Step 2. Goal G2.3 supports G2 through the systematic application of a *Shared Control* STPA method to uncover all UCAs and loss scenarios pertaining to shared control. The first claim G2.3.1 relates to defining the scope of the STPA analysis. For STPA analyses this means identifying the stakeholders (evidence [D]), their losses (evidence [E]), and the associated vehicle level hazards (evidence [F]). The second claim relates to the systematic application of STPA guide words to identify all UCAs relevant to shared control (evidence [G]). For context, C5 lists the STPA guide words used [96]. The final claim G2.3.3 is that by considering both *Classic* and *Shared Control* loss scenario types, all loss scenarios attributable to shared control (evidence [H]) are identified.

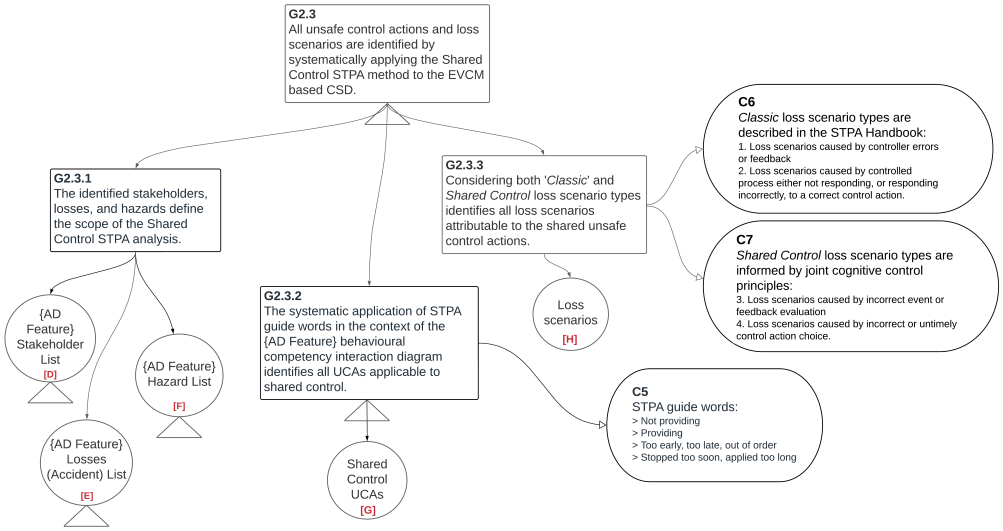


Figure 25: Goal 2.3: systematic application of a *shared control* STPA method assurance argument pattern

10.2.5 *Assurance Step 3: AD feature shared control specification assurance*

10.2.5.1 *Objectives*

Assurance Step 3 (shown pictorially in Figure 26) comprises two activities seeking to achieve the following objectives:

1. To define safety constraints (i.e. safety goals) that mitigate the identified hazards
2. To define safety requirements that mitigate all identified UCAs attributable to shared control and
3. To refine the safety requirements to mitigate all identified loss scenarios attributable to shared control.

10.2.5.2 *Inputs and outputs*

This assurance step elicits the safety requirements needed to adequately control system behaviour and interactions to mitigate hazards. Therefore, the key inputs to Assurance Step 3 are the hazard list and UCAs from Assurance

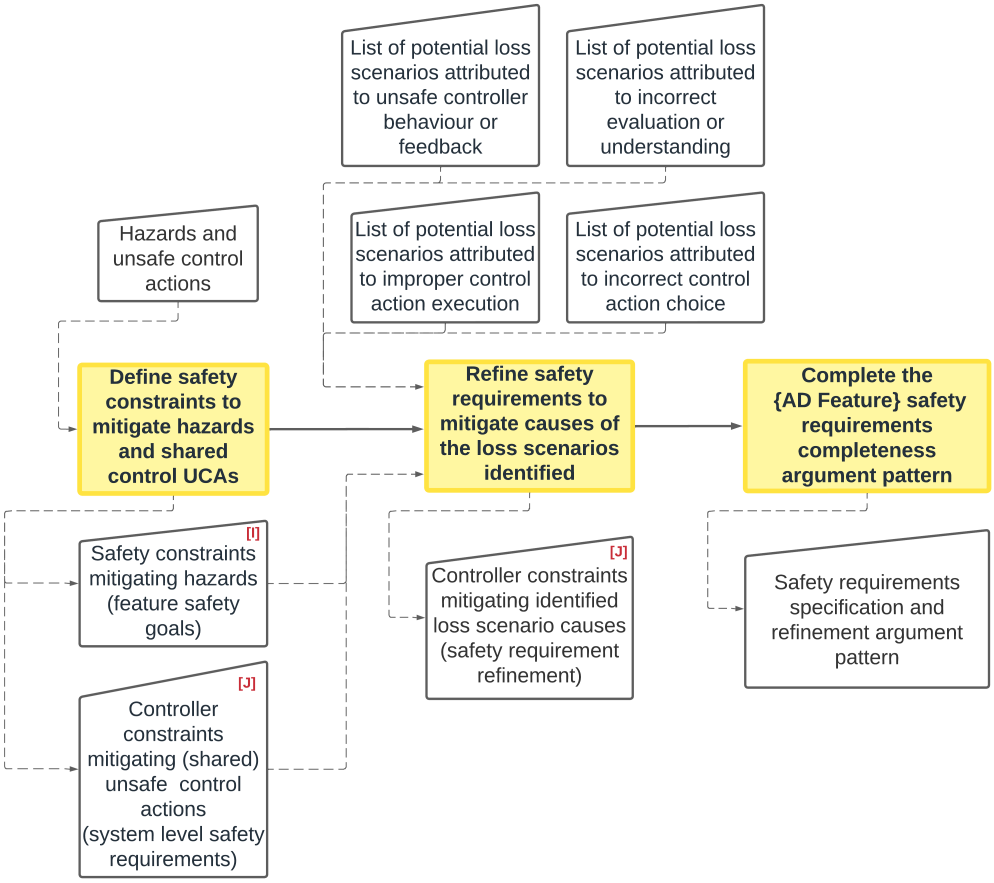


Figure 26: Assurance Step 3: {AD feature} shared control loss scenario mitigation process flow

Step 2. The safety constraints[3] for the {AD Feature} are derived from the hazard list, while the controller constraints[4] are derived to mitigate the UCAs previously identified. The safety requirements are then further refined, using the previously identified loss scenarios, to complete the safety requirements definition.

10.2.5.3 *Activities*

Assurance Step 3 comprises the following activities:

3 It is perhaps appropriate to consider safety constraints as being analogous to what ISO 26262 refers to as safety goals. Although safety constraints are typically broader in nature.

4 In a generic context, controller constraints can be thought of as safety requirements, at a level of abstraction below safety constraints.

Activity 1: Define safety constraints to mitigate hazards and shared control UCAs. The list of hazards and UCAs provides the input to this activity. Safety constraints are derived to mitigate the hazards, and the controller constraints to mitigate the UCAs that result because control is shared.

Activity 2: Refine safety requirements to mitigate causes of the loss scenarios identified. Not included within the *Classic* STPA process [96], but the loss scenarios identified provide a rich knowledge base from which to derive further safety requirements. For example, a potential hazard resulting from an incorrect control action selection could be mitigated with a safety requirement that assists the human agent to make correct and timely control action choices.

Activity 3: Instantiate {AD Feature} safety requirements completeness argument pattern. The final activity in Assurance Step 3 is to instantiate the safety case argument pattern given in Figure 27. The top-claim (goal G3) is that the safety requirements mitigate all shared control UCAs and loss scenarios identified. This claim is supported by three further goals: G3.1, G3.2, and G3.3. Goal G3.1 claims that the derived safety constraints (or safety goals) mitigate all identified hazards (evidence [I]), and goal G3.2 claims that the safety requirements written mitigate all identified shared UCAs. While goal G3.3 complements G3.1 and G3.2, by using the loss scenarios identified to further refine the safety requirements, with the assertion being that by considering all four loss scenario types (see Section 9.3.4), {AD Feature} safety requirements are identified that would otherwise have remained undiscovered (evidence [J]).

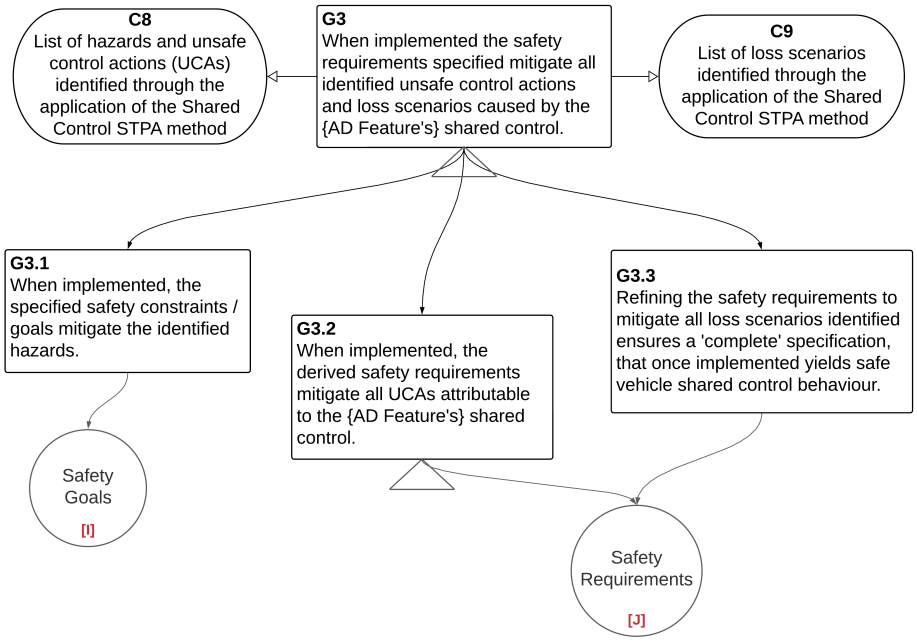


Figure 27: Goal 3: {AD feature} safety requirements definition assurance argument pattern

10.2.6 *Assurance Step 4: AD feature shared control is demonstrably safe assurance*

10.2.6.1 *Objectives*

Assurance Step 4 (shown pictorially in Figure 26) comprises three activities. These activities aim to demonstrate the safe shared control behaviour for the {AD Feature}, by seeking to fulfil the following objectives:

1. To define and conduct a test specification that demonstrates that the {AD Feature} correctly implements the safety requirements defined in Assurance Step 3 and
2. To devise a scenario based test strategy that demonstrates that the {AD Feature} behaviour is safe for all identified loss scenarios.

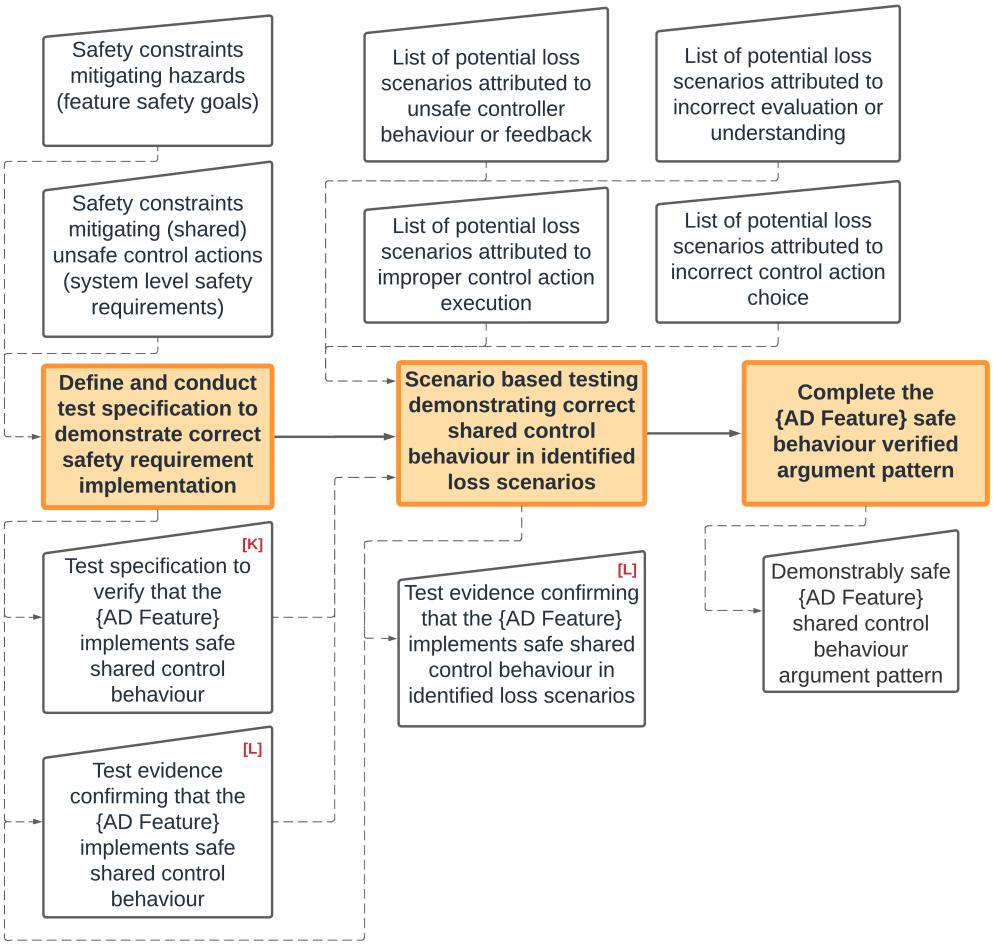


Figure 28: Assurance Step 4: {AD feature} shared control verification process flow

10.2.6.2 *Input and outputs*

The safety requirements specified during Assurance Step 3, together with the loss scenarios identified during Assurance Step 2, are the inputs to this step. The first activity in Assurance Step 4 is a traditional requirements verification task. That is, to write a series of test specifications for each safety requirement, and then to run those tests, demonstrating that the {AD Feature} satisfies its safety requirements. In an automotive context, test cases will typically be derived using the test derivations methods from ISO 26262-4:2018 Table 3 [162]. The final output is the test evidence from scenario based testing. Here, the loss scenarios identified during Assurance Step 2 form the test cases demonstrating that the {AD Feature} behaviour remains safe under all identified loss scenarios.

10.2.6.3 *Activities*

Assurance Step 4 comprises the following activities:

Activity 1: Derive and conduct test specification to demonstrate correct safety requirement implementation. Typified by a requirements led process, this activity derives the test specifications (Figure 28 artefact [K]) to verify that the {AD Feature} satisfies each of its safety requirements. The output from the test activity then produces the test evidence (artefact [L]) demonstrating that the {AD Feature} implements its safety requirements.

Activity 2: Scenario based testing demonstrating correct shared control behaviour In addition to the requirements based testing described in Activity 1, this activity uses the loss scenario catalogue to enrich the test activity. This increases confidence in the safety of shared control, by testing the {AD Feature} in scenarios where shared control has been identified as potentially hazardous. The loss scenarios identified during Assurance Step 2 provide the scenarios used by the scenario based testing to demonstrate that correct {AD Feature} behaviour is achieved (artefact [L]) for all shared control identified loss scenarios (artefact [H])[5].

Activity 3: Instantiate {AD Feature} safety verification argument patterns. The activities within Assurance Step 4 support goal G4 (see Figure 29), which claims that when tested the {AD Feature} exhibits safe shared control behaviour. G4 is then supported by two further sub-goals, G4.1 and G4.2, which make claims about the completeness of the test specification used. G4.1 claims that the use of *good practice* systems integration and testing derivation methods (e.g. ISO 26262-4:2018 Table 3 [162]) to derive test specifications, demonstrates correct implementation of the {AD Feature} safety requirements. Then G4.2 complements G4.1, by claiming that by using the identified loss scenarios catalogue, to

5 It should be noted that the shared control related loss scenarios identified as part of this process will typically represent a sub-set of all test cases. In practice, scenarios seeking to exercise particular corner-cases or triggering conditions (identified by SOTIF analyses) will also be included.

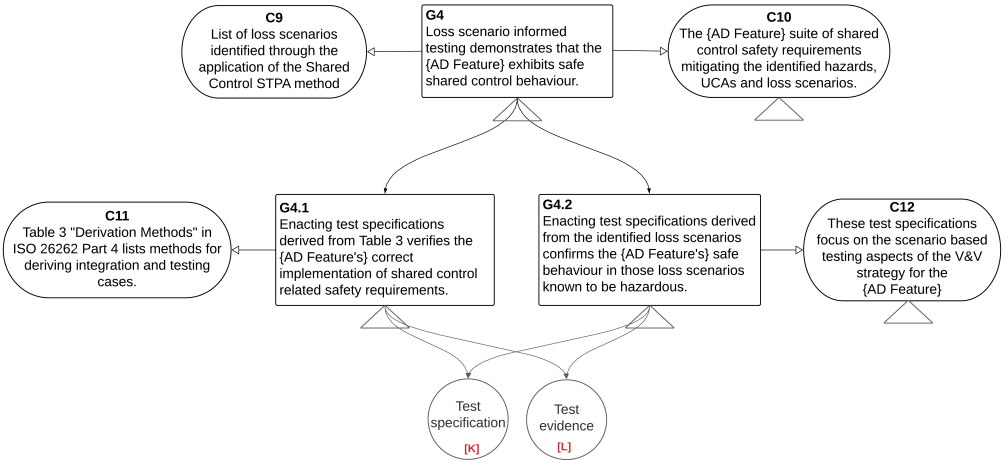


Figure 29: Goal 4: {AD feature} shared control behaviour is demonstrably safe argument pattern

derive further test scenarios, creates test evidence demonstrating safe {AD Feature} behaviour in the known potentially hazardous scenarios.

10.3 THE BROADER SAFETY CASE

The expectation is that the shared control safety case argument pattern presented above, would form part of a larger safety case argument for the AD vehicle feature. By conceptualising the cognitive interactions that take place between the automation and the human driver, while an automated feature is actively controlling the vehicle, a thorough examination of shared control hazards and hazard causes is possible.

Addressing shared control means this safety case argument directly supports argument claims about the safety of an AD feature's shared control behaviour while active and inside the ODD (i.e., *State 1* as described in [23]). In addition, potential hazard scenarios exist when control is being transferred between agents when transitioning into or from *State 1*. For example, the process of relinquishing control to the automation when the driver initially activates the AD feature, or the process of handing back control to the driver when the automation detects that the vehicle has left or is about to leave the ODD, or

a fault exists which means the AD needs the driver to resume control. Such control transferal brings with it inherent risks, so the inclusion of a robust safety case argument about the systematic analysis of shared control can only add confidence to AD feature's safety case.

10.4 SUMMARY

This chapter has described a method for developing a safety case argument structure for the shared control aspects of an AD vehicle feature. The expectation is that this shared control safety case argument will evolve in parallel with the feature development, and will support the safety case argument for the AD feature. The methodology follows a requirements led approach, typified by standards such as ISO 26262, but then uses the knowledge contained within the identified loss scenarios to enrich both the safety requirements elicitation and the test strategy derivation. This is achieved by enabling the safety analyst to conceptualise the nature of the AD feature's shared control. Although the method described here does not create the complete safety case argument, following this method does enable the safety practitioner to articulate a comprehensive argument about the safe implementation of the AD feature's shared control.

Part IV

CASE STUDY EVALUATION

The research activity has produced a hazard analysis method for shared control, comprising an Enhanced Vehicle Control Model, an analysis method that incorporates joint-cognition principles, and a safety argument pattern for shared control, which were described in Part III.

Part IV evaluates those research products using a case study evaluation method. With the case study evaluation comprising a number of cases, the evaluation strategy is introduced first in Chapter 11. Chapter 12 then discusses the evidence that supports the evaluation strategy. Chapter 12 concludes with a summary of the extent to which this evaluation supports the assertion that the research products do provide an *effective* hazard analysis method for shared control.

11

EVALUATION STRATEGY

With AD vehicle features challenging the functional safety paradigm (see Section 1.2.2 page 17), the aim of this research has been to contribute positively to automotive systems safety analysis. In doing so this research seeks to answer the question "how can the safety of AD be assured under different levels of shared human vehicle control?" The ambition of this research has been to provide a conceptual model and accompanying method that aids the safety analyst to conceptualise and explore shared control hazards causes. Having developed an EVCM that allows an AD vehicle feature to be modelled conceptually, and having used joint-cognition principles to further develop the *Classic* STPA process that uses the EVCM, the evaluation must then ascertain the extent to which when "used together the EVCM and accompanying method provides the system analyst with an *effective* tool for analysing automated driving features whose control is shared."

As highlighted by Pumphrey [130] the rigorous evaluation of new hazard analysis techniques is problematic. This is due in part to the lack of feedback (unless an accident occurs), project complexity and scale making back-to-back comparisons prohibitively expensive, and the subjective nature of hazard analysis making the reason of any identified differences impossible to discern. Consequently, it has been evident from the start that a combination of evaluation techniques would be needed to evaluate the *effectiveness* of the research products.

The evaluation evidence supporting the above proposition is drawn from a number of smaller case studies. The focus of these case studies has been to demonstrate the positive qualities (see Section 3.2) of the research products created. That is, the EVCM, behavioural competency taxonomy, shared control STPA method, and accompanying safety case argument pattern. In doing so, the

evaluation strategy has sought to establish these research products as a *proof-of-concept* method. However, further cases studies and trials would be needed to establish the viability of this method for use by design teams developing production AD vehicle features.

This chapter describes the strategy used to evaluate the various properties of the above research products, and seeks to argue that in combination these smaller cases evidence the research products as a *proof-of-concept* hazard analysis method for AD shared control. A GSN graphical representation of the evaluation strategy has been developed which appears in Appendix F page 308.

11.1 CONCEPTUAL MODELLING OF AD FEATURES

All hazard analysis methods require a conceptual model of the candidate system from which to begin. This research postulates that the EVCM is such a conceptual model; capable of representing the control structure of automated vehicle systems as a conceptual model[1]. So the first evaluation objective is to determine the extent to which the EVCM can represent vehicle AD features having differing SAE automation levels. AD vehicle features including ACC (SAE Level 1), HAS (SAE Level 2) and Automatic Lane Keeping System (ALKS) (SAE Level 3) have been used as candidate systems for this.

Given that the conceptual model is supporting a hazard analysis method, it is important that the feature functionality is also accurately described. As discussed in Section 8.2, in a manual driving context the automotive safety analyst will likely have a well honed mental model of *"driving"* – what contemporary research and standards (e.g. [141, 163, 173]) refer to as the DDT. Various interpretations of the tasks / behaviours needed to fulfil the DDT exist. However, these tend to focus on what the automation needs to achieve, rather than the tasks / behavioural competencies that the driver and automation need to achieve together to safely maintain vehicle control. This research presents a behavioural competency taxonomy (see Appendix A) for this task.

1 The EVCM has been evaluated in-part through a peer reviewed journal [108].

Consequently, an objective of this evaluation must be to determine whether the behavioural competency taxonomy can describe the task of driving, particularly when that task is shared between the human driver and the automation.

11.2 AN APPROPRIATE AND FEASIBLE CSD FOR STPA

This research proposes an enhanced version of *Classic* STPA as its hazard analysis method. Like other hazard analysis methods, critical to STPA's success is a conceptual model of the candidate system. In the *STPA Handbook* Leveson and Thomas discuss the importance of modelling at an appropriate level of abstraction when creating the CSD. However, as discussed in Section 5.2.1.1, the complexity of automotive system CSDs presented in the literature would suggest that in practice using abstraction to manage complexity is more difficult than the theory advocated by Leveson and Thomas would suggest.

To support the analyst to create an AD vehicle feature CSD, at a level of detail that supports the analysis of shared control, this research postulates that instead of the analyst creating the CSD from first principles, the EVCM should be their starting point. Then by using the EVCM in combination with the behavioural competency taxonomy, the AD vehicle feature can be described in sufficient detail to support the exploration of hazards caused by the AD feature's shared control.

The ability to use the EVCM as a CSD, and to use the behavioural competency taxonomy to describe that feature's behaviour, in sufficient detail to analysis shared control hazard causes has been evaluated in two parts. Firstly, the desk based analysis of ALKS [180] and HAS [6] seeks to demonstrate that used together the EVCM and behavioural competency taxonomy can describe AD vehicle features that represent industry "state of the art". The obvious criticism that should be levelled at such a desk based evaluation is that it is purely demonstrating the effectiveness of a conceptual model in the hands of its inventor! It is failing to demonstrate whether others can use the EVCM together with the behavioural competency taxonomy to model their own candidate

systems. To address this an STPA workshop has been run with a group of automotive design engineers from Oxbotica. Part of the workshop involved the participants using the EVCM to model shared control in the context of their Selenium product [120].

11.3 A SHARED CONTROL STPA METHOD

As described in Section 5.2.1.1, STPA is the hazard analysis method typically used by automotive practitioners to analyse AD vehicle features – particularly where system perception and hence SOTIF can affect system safety. Although the literature provides many examples of the successful analysis of AD systems using STPA, promoting the human in the system to the top of the CSD hierarchy has drawbacks (see Section 5.2.3). In contrast, considering the effects of shared control throughout a system's control hierarchy is unique to this research. This research postulates that modelling shared control in this way, better represents the true nature of driving shared control than does simply putting the driver at the top of the CSD. Therefore, an important aspect of the evaluation is to determine to what extent conceptualising shared control supports the analyst in their work. That is, to answer the specific question *"does the Shared Control STPA method identify more loss scenarios associated with shared control, which might otherwise have remained undiscovered?"* Using the EVCM and behavioural competencies to represent the candidate system supports an in-depth exploration of shared control. This is achieved by making explicit the applicable behavioural competencies, the responsibility for undertaking those competencies, and the potential interactions between competencies. The other aspect of the *Shared Control* STPA method is the introduction of two further loss scenario types. These are the loss scenario types associated with the joint cognitive aspects of shared control, relating to making an incorrect determination about an event as it happens, or incorrect or slow control action selection.

Again the evaluation of the *Shared Control* STPA method is undertaken in two parts. First the output of both the ALKS (Appendix D) and HAS (Appendix C)

vehicle features are compared to the available literature for these or similar systems. With this comparison focusing specifically on instances where using the *Shared Control* STPA method has uncovered potential loss scenarios associated with shared control not identified in the literature. The second part of the evaluation, based on an in-depth case study, draws on the Oxbotica Team's experiences using the *Shared Control* STPA method. That is, did the Team uncover anything about shared control hazard causes / loss scenarios associated with Selenium that had not been previously identified?

11.4 SHARED CONTROL SAFETY CASE ARGUMENT

The final research product is the shared control safety case argument pattern. This final part of the evaluation instantiates the safety case argument pattern for ALKS, to show how one might use the artefacts created by the shared control STPA method to present a safety case argument for shared control. The expectation is that the shared control safety case argument would complement a broader product safety case.

The safety case argument pattern includes the claims that both requirements elicitation and scenario based testing are enriched by the knowledge gained while identifying loss scenarios relevant to shared control. There is no reason to suggest that these claims are misplaced. However, the decision has been taken not to demonstrate this aspect as part of the evaluation. With both requirements elicitation and verification and validation strategies being huge topics in their own right, they are better addressed by researchers with expansive experience in those fields.

12

EVALUATION EVIDENCE

Using the evaluation strategy described in Chapter 11 this chapter discusses the evidential material supporting the evaluation. In doing so, the discussion seeks to highlight where the evaluation evidence either supports or opposes the research products properties introduced in Section 3.2. As well as demonstrating that when used together the EVCM and *Shared Control* STPA method are *effective* at analysing shared control hazard causes, this discussion must also provide a candid assessment of the contribution that can be claimed. As such, it must highlight when the evidence is inconclusive or could be interpreted differently.

This chapter is structured similarly to the previous chapter discussing the evaluation strategy, and like that chapter, is structured following the GSN argument presented in Appendix F on page 308.

12.1 CONCEPTUAL MODELLING OF AD FEATURES

Two formats of the EVCM have been evaluated, that is a control system style version (as depicted in Figure 14, page 115), together with a hierarchical controller style version for use as the CSD in an STPA based hazard analysis. For example the CSDs created for HAS (Figure 33, Appendix C, page 261) and ALKS (Figure 36, Appendix D, page 284) evaluation case studies.

Evidence from the desk based case studies[1] support the premise that AD vehicle features having different levels of automation can be modelled conceptually using the EVCM [108]. In addition, when the EVCM is then used in conjunction

1 Specifically, ACC (SAE Level 1), ACC with Lane Centring (SAE Level 2), and TJA (SAE Level 3) in [108], and HAS (SAE Level 2) and ALKS (SAE Level 3) in Appendix C and Appendix D respectively.

with the behavioural competency taxonomy (see Appendix A, page 210) an AD feature's control, manoeuvring and strategic level behaviour can be described.

During the Oxbotica workshop (see Appendix E page 298) the team members responsible for the Selenium automated design were supportive of the inclusion of the control hierarchy in the EVCM. With Oxbotica Team members suggesting that by considering the different levels of a system's behaviour (i.e. control, manoeuvring and strategic), the analysis process aligns with the way in which engineering teams typically consider a highly automated system's behaviour. The Oxbotica team felt that this is particularly true for machine learning implemented manoeuvring level type functions, where complex functionality lends itself to being described in a linguistically natural way [N. Ratiu, personal communications, 3rd November 2021].

As discussed in Section 2.1.3 (page 29) the behavioural competency taxonomy is based on the HTAoD – a taxonomy first developed to describe the control, manoeuvring and strategic level tasks associated with controlling a manual transmission passenger vehicle [186]. Modelling driver assistance (SAE Level 1) and partial automation (SAE Level 2) vehicle features, during the course of this research, has highlighted that the HTAoD is not an exhaustive task list in situations where the driver is expected to interact with automation. Research suggests that using automation adds to a driver's workload and potentially introduces new tasks to be performed [17, 19, 149] . Therefore, it is perhaps unsurprising that this evaluation has identified tasks, not included in the original HTAoD, that the driver also needs to fulfil when using vehicle automation. For example, understanding the capability and limitations of a given AD vehicle feature and *use automation appropriately*, or continuing to *monitor automation* while the vehicle feature is operational – again to ensure the automation operation remains safe in the presence of numerous external factors.

The ALKS evaluation study (see Appendix D page 272) begun by coding the requirements of Regulation 157 [180] into the relevant behavioural competencies using the NVivo qualitative analysis tool. The output from this

activity generated two "Sunburst" diagrams (see Figure 34, page 274) – the first "sunburst" (a) shows the behavioural competencies identified in a simple symmetrical pattern, while the second (b) indicates the proportion of requirements covering each behavioural competency. At a superficial level one might claim that as evidence the "sunbursts" support the premise that the behavioural competency taxonomy provides an *effective* vocabulary from which to describe the behaviour of a contemporary vehicle feature like ALKS. A more considered review of the "sunburst" diagrams (particularly (b)) highlights further observations that can be made about ALKS in relation to the behavioural competencies, namely:

- for an SAE Level 3 conditional automation feature, like ALKS, the majority of the relevant behavioural competencies identified are at the manoeuvring level.
- for ALKS the majority of behavioural competencies that require the human driver and the automation to share control happen at the manoeuvring level and relate to the behavioural competencies involving tactical decision making
- given the proportion of requirements in Regulation 157 relating to the *avoid obstacles* and *transition of control* the risks associated with the vehicle hitting another road user / object or the responsibility for vehicle control not being correctly transferred between the human driver and the automation are clearly behavioural competencies prominent in the Regulator's minds.

Feedback during the Oxbotica workshop (see Appendix E) indicates that the need for additional behavioural competencies is not limited to the lower automation levels. For the Lingen BP Refinery example [26] the human operator will typically be remote to the vehicle, highlighting that the EVCM and behavioural competency taxonomy may require expanding to consider complex human and organisational factors. However, such organisational considerations are outside the scope of this research. For an SAE Level 4 system like Selenium the team suggested that the system also needs to implement functionality to manage the *transfer of control* between the human operator and the automation.

The authors own analysis of HAS and ALKS (see Appendix C and Appendix D respectively) supports this view. With partial automation systems such as HAS needing to implement functionality to assess *driver engagement* and to manage the *transition of control*, while ALKS identified the need for *ODD detection* functionality, to ensure that the automation can only operate while the vehicle is within the ODD.

As stated above, this research postulates that shared control typically takes place when the system is undertaking manoeuvring level behavioural competencies. During the Oxbotica workshop, the general discussion on the behavioural competency taxonomy and actor responsibility identification activity also identified that shared control typically occurs at the manoeuvring level. Thus corroborating the author's own findings.

The Oxbotica design team observed that by describing shared control in the context of behavioural competencies and considering human control throughout, the EVCM "helps keep the human inside the control loop and inside the system", allowing shared control to be "considered more completely" than is possible when the human driver is modelled at the top of the CSD. In addition, the observation was made, that when complete, the behavioural competency responsibility table[2] made explicit those behavioural competencies implemented by the system – and perhaps as importantly, those behavioural competencies not implemented by the system – and as such would make a useful addition to the Item Definition[3].

The behavioural competency taxonomy has been introduced within the MISRA Safety Case Working Group where work to develop safety case argument patterns for highly automated driving applications is ongoing. Here the observation is made that when industry considers AD vehicle feature behaviour, it is typical to focus on the behaviours that comprise the DDT. However, if the driver can be removed from the vehicle control loop for periods of time,

2 For commercial reasons it has not been possible to include the behavioural competencies responsibility tables from the Oxbotica workshop in this thesis. For similar behavioural competency responsibility tables created for HAS and ALKS vehicle features please see Table 9 Appendix C page 257 and Table 16 Appendix D page 279 respectively.

3 The Item Definition is a work product required by ISO 26262 [162].

then not only must the automation deliver the DDT while it is active, but it must also deliver the vehicle oversight tasks typically undertaken by the driver. The MISRA Safety Case Working Group members suggested that the behavioural competency taxonomy provides a means by which these broader *driving enterprise*[4] behaviours, needed to successfully navigate a vehicle within the road network, can be expressed and reasoned about [MISRA Safety Case Working Group 2 day workshop, 29th / 30th March 2022].

Having introduced both the MISRA Safety Case Working Group and the Oxbotica team to the behavioural competency taxonomy and observing the groups' interaction with the concepts, a functional safety expert present at both events said that "this research's impact on the thinking of those automotive safety practitioners should not be underestimated." Suggesting that by using the concepts of control, manoeuvring and strategic hierarchical levels of control has the effect of broadening people's thinking – particularly compared to functional safety concepts and techniques that tend to only focus thinking at the control level [N. Ratiu, personal communications, 3rd November 2021].

It should be noted that the behavioural competency taxonomy presented here is not an exhaustive list, and it is focused towards achieving ego vehicle safety. The case study evaluation has generated evidence that supports the view that the taxonomy adequately describes the behavioural competencies needed for ego vehicle safety. What the evaluation has not been designed to do is to uncover taxonomy limitations or to test the taxonomy beyond a single vehicle. For example, the behavioural competency taxonomy cannot describe the behavioural competencies needed by other actors, such as cyclists or pedestrians, nor can it describe behavioural competencies needed to communicate with or to understand the intentions of other road users. A quality that phycology research refers to as the theory of mind [195]. However, the case studies undertaken by this research have shown the taxonomy to be expandable. For example, the analysis thus far has not considered any vehicle AD features that implement functionality to protect the rear of vehicle. However,

4 *Driving enterprise* is a term devised by the MISRA group to express this wider notion of driving. For example, checking that the vehicle is in fully working order before beginning a journey, or anticipating the behaviour of other road users during the journey.

were such an AD feature being analysed then this could be addressed with the addition of a behavioural competency, such as *maintain distance to the succeeding vehicle*.

The EVCM has also been used to conceptualise AD in different innovative ways. As part of an ongoing HORIBA MIRA commercial project, a colleague is working with a vehicle manufacturer to develop a data model of vehicle and item level hazards. The EVCM and the blocks that comprise the EVCM (e.g. strategy level, manoeuvring level, models / perception) are providing the structure for the data model. This allows relations between vehicle level movement constraint [132] violations and item level hazards (e.g. drive torque greater than demand) to be made and reasoned about - thus, supporting future AD feature development [J. Birch, personal communications, 26th July 2022].

12.2 AN EFFICIENT CSD FOR STPA

Incorporating a hierarchical control structure (i.e. the control, manoeuvring and strategic levels) means the EVCM can easily be redrawn as a hierarchy of controllers – thus making it analogous to a CSD used in STPA. In Appendix C this research demonstrates using the CSD version of the EVCM to model a partial automation HAS vehicle feature (incorporating lane centring and ACC functionality). In this example the HAS control behaviour is described using 12 behavioural competencies – two of which are at a strategic level, five are at a manoeuvring level and five are at a control level. The interaction of these 12 behavioural competencies and their control of the vehicle plant is then described by 9 control actions.

However, creating a behavioural competency responsibility matrix (see Table 9) and behavioural competency interaction diagram (see Figure 32) for HAS, highlights that shared control is taking place between three behavioural competencies. Those behavioural competencies are *maintain position in lane, follow other vehicles* and *maintain speed*, and important to those behavioural

competencies are the control actions *target trajectory* and *longitudinal acceleration*. Therefore, the analyst is able to focus their effort towards the two control actions being influenced by shared control. Modelling HAS in this way allows the analyst to consider the system in sufficient detail to facilitate identifying hazard causes, while not becoming overwhelmed by the control structure detail. In contrast, the CSD created by Becker et al. to describe only lane centring functionality [see 21, page 29] is complex in comparison – with Becker et al.'s CSD including over 20 individual control actions. In addition, a description of the lane centring behaviour is not provided in [21]. If one considers that the STPA method requires the analyst to apply multiple guide words to each control action, it is evident that Becker et al.'s complex CSD requires analysing approximately 60% more control actions than the HAS example presented here (see Appendix C), and potentially not find additional hazard causes.

In addition to providing a framework in which to describe the control action hierarchy, the HAS and ALKS evaluations (see Appendix C and Appendix D respectively) demonstrate how the EVCM's framework facilitates the identification of input variables needed by each control level. For example, the raw vehicle speed and time-of-flight sensor data inputs to the control level (see Figure 33, page 261) and the pre-processed (by the control level) object position and classification information inputs to the manoeuvring level. This is in contrast to Becker et al.'s CSD that focuses on raw signal data (e.g. vehicle lateral acceleration, lane boundary information); where little or no inference is made by the system about the environment in which the vehicle is operating (i.e. situation awareness). With the EVCM including the physical attributes relating to the operating environment (physical, traffic, environmental), the potential for SOTIF hazard causes remains prominent. For example, prompting the analyst to consider the impact of environmental factors, such as ambient temperature or traffic density, on both the behaviour of the system and its interaction with the human driver (i.e. the effect on shared control); with Stanton suggesting that often it is difficult to consider environmental factors with *Classic* STPA [148].

For instance, this research has identified that for both HAS and ALKS, temporary construction zones (e.g. road works) include environmental factors (i.e., SOTIF triggering conditions) capable of causing a hazard. For example, when the automation's lane centring behaviour continues to follow the path depicted by the old erased road markings rather than the new temporary road markings. No comparable observation was found in [21]. Whether it is the use of the EVCM and behavioural competency taxonomy that has facilitated finding these potential environmental triggers, or whether it is the author's epistemic knowledge and experience that has made such observations possible is impossible to tell. However, minimising the CSD's complexity does undoubtedly help the analyst avoid being distracted by low-level abstraction detail before the analysis has been completed for the higher levels of abstraction. In addition, using the behavioural competency taxonomy certainly provides the analyst with a useful *aide memoir* and experience suggests that analysing the whole system at a high level of abstraction first, before then using the results of that initial analysis to inform more focused STPA iterations is beneficial [D. Higham, personal communications, 17th April 2020].

The importance of using abstraction to manage complexity is well documented (e.g., Section 5.2.1.1), and this research asserts that by using the EVCM as a CSD, to inform shared control vehicle level hazard analysis, the following benefits are realised: Firstly, external influences on the AD vehicle feature (e.g., environmental factors) can be modelled. Secondly, inferences made within the system (e.g., object position, speed and classification, or the intention of other road users) can be modelled, and thirdly, by virtue of modelling a system with fewer UCAs than is typically seen (e.g., the CSD presented by Becker et al.) makes the analysis shorter and thus more efficient. Although the evidence from the case studies undertaken thus far, does not refute the above suggested benefits, it does not positively support the claim either. Further case studies, ideally undertaken on a production intent AD vehicle feature, would be needed to compare a CSD created from first principles to that created using the EVCM as its starting point. Then the comparative effort required could be determined, together with a comparison of the information captured by each model. Each analysis team would then need to complete the remaining STPA steps and

the outputs from the two sets of analyses compared. This would determine whether the two cases identify similar hazards and hazard causes; with the caveat being that some differences are to be expected given the subjective nature of the hazard analyses process.

12.3 THE SHARED CONTROL STPA METHOD

12.3.1 *Analysis of a lane centring function*

To test the effectiveness of the EVCM with the *Shared Control* STPA method, Becker et al.'s published analysis of a lane centring system has been used as a comparison [21]. Becker et al. have used a combination of HAZOP, functional failure analysis and STPA to identify system hazards, to assess their risk, identify hazard causes, define safety concepts and to describe those safety concepts as safety requirements.

A key assertion made by this research is that the EVCM is capable of modelling automated vehicle features that include shared control, at a level of abstraction necessary to perform hazard analysis. However, attempting to model Becker et al.'s system [21, figure 6, page 29] using EVCM did highlight that detailed subsystem interactions are less readily conceptualised using the EVCM. The specific subsystem interaction identified relates to achieving the desired lateral vehicle control through the application of asymmetric wheel torque.

Achieving the desired lateral control functionality relies on what Becker et al. refers to as "foundational vehicle systems" interacting. That is, the vehicle's steering, braking and lateral control, propulsion, and active differential systems. For example, the primary actuation means might be the steering system applying steering torque directly to the vehicle's front wheels. However, there may be operational or "fall-back" situations when the asymmetric torque generated by the active differential system and / or the brake / stability control systems are used to achieve the desired vehicle yaw moment. Yaw control such as this requires *yaw stability coordination* messages to pass between

the three sub-systems. It is not possible to model this coordination detail explicitly using the EVCM. That said, an error in *yaw stability coordination* would ultimately result in the steering torque control action, and the subsequent Vehicle System Behaviour being wrong, which could be implicitly modelled. Also, the interaction between the asymmetric torque generating sub-systems could be explored during STPA Step 4 – loss scenario identification.

Related to the above discussion about *yaw stability coordination* is a surprising hazard omission from [21]. That is, in the lane centring study by Becker et al., no consideration of loss of vehicle stability, as a potential vehicle level hazard, was found. Given that ALC directly controls vehicle steering, and hence vehicle yaw, not including the loss of vehicle stability as a hazard is a critical omission.

Modelling driver assistance and partial automation systems (specifically ACC and ACC with lane centring) in [108] emphasised the stark differences that may exist between the information the human driver takes from their environment compared to the information that the automated system is probably able to infer from the same scene. For example, the automation may only have time-of-flight RADAR data from which to model the behaviour of other road users, while the human driver may be using sight to judge distance, but also to gain valuable visual cues. These differences are potentially most acute when the scenario involves interactions with other actors – i.e. other road users. Considering this difference in perception between the human driver and the automation raised further questions about the impact this might have on vehicle safety, particularly in relation to the transfer of control and shared control in general. Hollnagel's cyclical models of human action (i.e. COCOM and ECOM) provided a concept and STPA the method from which to reason about shared control.

Using the *Shared Control* STPA method to analysis the HAS vehicle feature, focused the analysis on two manoeuvring level control actions, namely: *target trajectory* and *longitudinal acceleration / deceleration (target speed)*. Following the process identifies UCAs that arguably one would expect to identify using *Classic* STPA. For example, the driver electing to use the automated vehicle feature when environmental conditions (perhaps snow is falling or there is thick fog)

mean that it is inadvisable and potentially unsafe to do so. However, following the *Shared Control* STPA method with the CSD version of the EVCM also identifies potential unsafe control actions relating to shared control in the context of specific behavioural competencies. For example, during an evasive manoeuvre the driver makes a corrective input, but the steering input they make is too small. The reason for this might be that the driver has been out-of-the-loop for an extended period, or during the steering intervention the driver stopped applying a steering input because they felt the automation acting against them.

The motivation for using the research from [21] as part of this evaluation was to compare the STPA analysis outcomes from Becker et al.'s research with the findings from this research, using the *Shared Control* STPA method. With the ambition being to demonstrate how the *Shared Control* STPA method identifies loss scenarios not discovered by the former analysis. However, having reviewed the output from the HAS case study with the work of Becker et al., it was clear that such a comparison was not possible.

Becker et al. limited their analysis to functional safety considerations, stating that *"additional analysis is necessary to identify safety considerations that do not result from malfunctioning electronics (e.g., safety of the intended function)"* [21, page 14]. The intention of the *Shared Control* STPA method is that its focus is broader than functional safety, particularly in relation to shared control, which by its very nature means that hazard causes will likely be due to differences in perception between the human driver and the automation. Becker et al.'s research also considers the function of lane centring (a generic function applicable to different levels of automation) in isolation[5], and does not appear to apply the full STPA method to that function. In contrast, using HAS as the example, gives a tangible vehicle feature to reason about, together with the following aspects has led to analysis outcomes that are difficult to compare: Using the EVCM derived CSD allows the analyst to reason about the system at a

5 In undertaking this evaluation I found it difficult to reason about lane centring in isolation, so have some sympathy with the work undertaken [21]. Adding longitudinal control and being able to use the *Shared Control* STPA method to analyse a real vehicle feature, like HAS, has facilitated a more complete and detailed *thought experiment*.

higher level of abstraction and in a broader context. The assertion here is that this makes the process efficient, which makes the application of the whole STPA process easier and consequently achievable. So, with outcomes from the latter analysis (see Appendix C) not being readily comparable with the prior research [21] leads to the above admission that it has not been possible to compare like with like.

12.3.2 *Analysis of a vehicle regulation*

Using the *Shared Control* STPA method (see Chapter 9 page 123) to analyse ALKS (see Appendix D) demonstrates the methods capability to analyse a contemporary vehicle feature; whose safe operation relies upon correct shared control. This observation about the importance of shared control to ALKS is borne out by the number of requirements in the ALKS Regulation [180] relating to the topics of driver awareness and the transfer of control. With the regulation including at least 5 clauses relating to driver availability and awareness and 19 clauses addressing the transition of control.

Like the HAS example discussed above, it is while undertaking *manoeuvring* and *strategic* behavioural competencies that vehicle control is shared between the human driver and the automation. Specifically in relation to *performing surveillance*, *respond to traffic conditions* and *navigate temporary conditions*, which requires the human driver to provide oversight of the transfer of control management performed by the automation. The applicable control actions relating to the transfer of control between the human driver and the automation being: *cancel automation* and *driver transition of control request*. While the unsafe control actions identified relating to the incorrect timing of *driver transition of control requests* – for example, the driver is not afforded sufficient time to successfully resume control, or the driver is not actually back in the control loop when the automation ceases DDT control. Applying the joint cognitive informed loss scenario identification questions (see shared control questions 3 and 4, Figure 17, page 137) helped identify loss scenarios where timing is important and where transition of control success might be affected by the situation in

which the transfer of control happens. For example, the relative success of a take-back control request happening while driving on a quiet straight road would be very different to the same take-back control request taking place on a busy narrow multi-lane commuter road. Also if the exact point at which the automation releases DDT control is not clearly communicated and evident to the driver then they might fail to make the required and timely control inputs. Having identified such potential transfer of control shortfalls, safety requirements are written to mitigate the loss scenarios identified (see Table 23). For example, "R25: ALKS related messages shall clearly communicate the next action that the driver must take", and "R27: ALKS transfer of control duration shall be increased to account for environmental conditions detected".

12.3.3 *Analysis of Selenium*

Although the Oxbotica workshop (see Appendix E page 298) introduced the Oxbotica engineering team to the EVCM and the *Shared Control* STPA method, the workshop duration was inadequate to give the design team sufficient time to apply the principles to the Selenium system and to provide critical feedback on the *Shared Control* STPA method. However, having seen how the CSD version of the EVCM can describe an SAE Level 4 system like Selenium, at a level of abstraction applicable to AD systems incorporating machine learning, the Oxbotica team intend to use the *Shared Control* STPA method to complement the other safety analysis techniques currently used within the business. This was seen as advantageous compared to functional safety techniques, such as FTA, that the Oxbotica team suggested tend to be "too low a level of detail for AI systems". The expectation is that by aligning the system architecture to the control, manoeuvring and strategy hierarchical levels will make the safety requirements elicitation process efficient by allowing safety requirements identified during Step 4 of the process to be directly allocated to the appropriate control level in the system architecture.

12.3.4 *Further evaluation thoughts*

As discussed above, the *Shared Control* STPA method evaluation has proved somewhat inconclusive, particularly with regard to supporting the assertion that by

> focusing on behavioural competencies with shared actor responsibility and using cognitive principles to enhance *Classic* STPA loss scenario types, allows the EVCM and accompanying method to identify hazard causes and hazardous situations associated with shared control that might otherwise remain undiscovered.

To test this assertion requires further case studies to be carried out, ideally on a production representative AD vehicle feature rather than the somewhat theoretical AD vehicle features evaluated so far. The analysis work undertaken within these future case studies should be done by the system design team members, with the case study researcher having less direct involvement in the analysis work than has be possible thus far.

The case studies undertaken here have considered the STPA method in its entirety, that is from Step 1 "Define purpose of the analysis" through to Step 4 "Identify loss scenarios". With the benefit of hindsight, creating a Logic Model [204, page 186] for both the *Classic* STPA and *Shared Control* STPA methods could identify individual process steps or artefacts where positive or negative change could be readily assessed. For example, in one case observing the team using the loss scenario questions (see Figure 17 page 137) to identify potential loss scenarios, while in the other case observing the team using only the material provided in the *STPA Handbook*. Using a Logic Model in this way should generate case study evidence that is more insightful that the case studies completed thus far, particularly regarding the relative merits of the *Shared Control* STPA method in relation to *Classic* STPA.

12.4 SHARED CONTROL SAFETY CASE ARGUMENT

The final research product, created as part of this research, is a safety case argument pattern for shared control. The safety case argument pattern complements the *Shared Control* STPA method, by showing how the evidence created while following the method supports an argument that an AD vehicle feature's shared control has been addressed adequately. The intention is that, once instantiated, the argument pattern would form part of a larger safety case argument for the AD vehicle feature.

The ALKS example (see Appendix D, page 272) demonstrates how a safety case argument for shared control can be constructed from the evidence created by the *Shared Control* STPA method. The safety case argument pattern for shared control includes an argument based on test evidence. However, with scenario based testing being outside the scope of this research ALKS related test evidence does not form part of this evaluation. That said, the ability to create a test scenario library using the loss scenarios discovered during the analysis is evident in the ALKS case study (see Tables 24 and 25 page 291). For example, the transfer of control between the driver and the automation could be unsuccessful for scenarios where the ego vehicle's lane ends, poor visibility affects ALKS perception system performance, or the driver is distracted and slow to regain control. Writing test cases for the loss scenarios identified will help enrich the verification programme.

The material in Chapter 10 has been shared with Oxbotica's Principal Validation Engineer, who felt that the overall safety case argument pattern was representative of a typical highly automated driving validation strategy. However, considering the tasks relating to the definition and satisfaction of non-shared control safety requirements, from those specific to shared control was felt to be somewhat artificial. Typically these activities are undertaken together, and "we'd use the scenarios to demonstrate correct implementation." Although the testing argument pattern is not reflective of how a testing strategy is typically created, making such an artificial distinction does ensure that the topic of

shared control remains explicit in the argument [A. Beaven, personal email communications, 24th November 2021].

It is understood that the testing tactic deployed will typically decompose each scenario into multiple subtests, with each subtest addressing specific aspects of system behaviour, or specific corner cases. For example, in a scenario that involves a cyclist, the test strategy might include tests to confirm that the system's perception system correctly detects the cyclist, identifies the cyclist as a cyclist, and tracks the cyclist's path correctly. Then given a defined path for the cyclist, the test strategy will also test that the system's path planning functionality devises a plausible path giving the location / proximity of the cyclist, and finally that the system's vehicle control functionality executes that plan correctly. From the feedback received it is clear that the safety case argument pattern for testing (see Figure 29, page 160) does not currently include sufficient detail to reflect the intricacies of an industrial scenario based testing strategy. However, goal G4.2 could be expanded to include a strategy that uses the loss scenarios identified during the *Shared Control* STPA method to inform the development of these highly structured test cases [A. Beaven, personal email communications, 24th November 2021].

12.5 SUMMARY

This research has sought to answer the question "How can the safety of AD be assured under different levels of shared human-vehicle control?" by creating a conceptual vehicle model, a behavioural competency taxonomy, a hazard analysis method, and a safety case argument pattern. When used in combination these research products form a concrete method from which to reason about AD vehicle feature shared control hazard causes. To evaluate the *effectiveness* of the method a series of case studies have been undertaken to establish whether the above products exhibit the following properties:

PROPERTY 1: The EVCM supports the conceptual modelling of AD vehicle features having shared control. And as a conceptual model the EVCM can be used to facilitate AD hazard analysis *thought experiments*.

PROPERTY 2: When used together, the EVCM and behavioural competency taxonomy supports AD hazard analysis, by providing a conceptual vehicle control model for AD vehicle features that include shared control.

PROPERTY 3: Focusing on behavioural competencies with shared actor responsibility and using cognitive principles to enhance *Classic* STPA loss scenario types, allows the EVCM and accompanying *Shared Control* STPA method to identify hazard causes and hazardous situations associated with shared control that might otherwise remain undiscovered.

PROPERTY 4: The supporting safety case argument methodology highlights potential loss scenarios attributable to shared control and emphasises how AD feature design modifications mitigate such hazard causes.

The extent to which the evaluation affirms or refutes the above properties is summarised below.

12.5.1 *Evaluating property 1*

The case studies have shown that the EVCM can be used to create a conceptual control model of an AD vehicle feature, with levels of automation from SAE Level 1 through to SAE Level 4 having been modelled. This research has focused on using the EVCM as a hazard cause identification aid, to model the interactions between the driver and the automation, and between the automation and vehicle's environment. A colleague has used the EVCM in a similar way. However, rather than using the EVCM as part of a hazard analysis method, they have used the EVCM elements as scaffolding, to manage the interaction between system and vehicle level hazards. The EVCM has not been evaluated outside of this hazard analysis context.

12.5.2 *Evaluating property 2*

Three case studies (i.e., HAS, ALKS and Selenium) use a hierarchical control structure version of the EVCM, which is analogous to the CSD needed by the STPA method. When used with the behavioural competency taxonomy these cases demonstrate how relevant control, manoeuvring and strategic behavioural competencies can be added to the EVCM, together with the relevant control actions (e.g., target trajectory, longitudinal acceleration), raw data (e.g., wheel speed, time of flight data) and information inferences (e.g., traffic density). Using the EVCM and behavioural competency taxonomy together in this way has proven successful for the cases undertaken. However, it should be noted that these case studies were motivated by the desire to demonstrate a shared control hazard analysis proof-of-concept. Therefore, the bias has undoubtable been towards demonstrating where the method works rather than identifying where it does not.

The use of the behavioural competency taxonomy, to describe the driving competencies needed to operate a vehicle in the ODD and to maintain safety, has shown how the human-machine interaction (i.e., the shared control responsibility split) between the human driver and the automation can be explicitly stated. Although the behavioural competency taxonomy seeks to describe all the competencies necessary to drive a passenger vehicle safely, it does not claim to be a complete list. Indeed, the case studies did identify missing competencies, such as *use automation appropriately* and *transition of control*, which have subsequently been added to the taxonomy that appears in Appendix A.

The case studies undertaken to date have not tested the taxonomy beyond a single passenger vehicle (e.g., to include pedestrians and vulnerable road users) nor have specific cases been designed to actively look for missing or incorrect behavioural competencies. Further case studies, undertaken by others and covering disparate real-world driving scenarios, would be needed to test whether the existing taxonomy is complete and correct.

12.5.3 *Evaluation property 3*

The evaluation's ability to support or refute Property 3 is mixed. The evaluation evidence thus far suggests the behavioural competency responsibility matrix to be a suitable mechanism to capture the responsibility for doing, monitoring and maintaining the safety of each applicable behavioural competency. However, when the Oxbotica team members completed a behavioural competency responsibility matrix for Selenium, the individuals' matrices were different. From the evaluations undertaken thus far, it is difficult to discern whether the differences found are due to a lack of method robustness, or due to system knowledge differences amongst the design team.

Testing the assertion that this method identifies shared control hazard causes that would otherwise have remained undiscovered was problematic. By comparing the output from the HAS case study with the lane centring analysis undertaken by Becker et al., the expectation had been to test this assertion. Differences in the hazard causes identified by the two analyses did exist, but it was not possible to determine the reason for those differences. Analysing the behaviour of a complete vehicle feature rather than a generic lane centring function, differences in abstraction levels between the two CSDs, the positive rejection of SOTIF related hazard causes by lane centring analysis team, likely differences in tacit knowledge between the analysts, and the *Shared Control* STPA method itself, all potentially contributed to the differences seen.

12.5.4 *Evaluation property 4*

The ALKS case study has demonstrated instantiating the safety case argument pattern, with the outputs generated by the *Shared Control* STPA method forming the evidence for that argument. The safety case argument pattern has also been reviewed by the Principal Validation Engineer at Oxbotica. They considered the safety case argument pattern structure to be representative of a highly automated driving validation strategy, but did feel that making the distinction

between shared control and non-shared control safety requirements somewhat artificial. Although, the ALKS example has shown how the loss scenario catalogue might inform the validation strategy for an AD vehicle feature, a targeted validation case study would be needed to establish the positive merits of this approach.

Part V

DISCUSSION AND CONCLUSION

Having completed this research journey, what observations can be made in relation to the literature, industry practice and the original research question? Specifically, how can the safety of AD be assured under different levels of shared human-vehicle control? This part concludes the thesis by reflecting on the research undertaken in the context of the problem definition, the literature, automotive state-of-practice and the original research question.

Chapter 13 reflects on the research undertaken in the context of both the 'state of the literature', and 'the state of practice', and where appropriate, observations about opportunities for future research activities are made. Then Chapter 14 makes the final concluding remarks about this research journey and invites others interested in shared control safety to pick up the Enhanced Vehicle Control Model (EVCM) and accompanying Method and to run with it.

13

DISCUSSION

This chapter draws the thesis to a close by reflecting on the research undertaken in the context of the problem definition, the literature, automotive state of practice and the original research question. The discussion reflects on the research undertaken in the context of both the 'state of the literature', 'the state of practice' – i.e. in relation to requirements of the applicable standards. – and the aims and objectives of this research. Where appropriate, observations about opportunities for future research activities, that relate to the EVCM, behavioural competency taxonomy and *Shared Control* STPA method, are made.

13.1 THE DRIVER AND SAFETY

The passenger car, the task of driving and road safety have all been evolving over the past 100 years or so. Early road safety initiatives saw mechanical systems, such as windscreen wipers, laminated windscreens and speedometers being fitted to passenger vehicles from the 1920's. The 1960's saw the introduction of global frameworks to harmonise vehicle performance, vehicle systems, conformity of production and in service inspections requirements. Then since the late 1990's consumer driven programmes, such as Euro NCAP *5 Star* rating, have sought to further improve road safety, which has undoubtably led to the accelerated introduction of AD vehicle features today.

Unsurprisingly the driver has had a prominent role in the story of road safety; being cast as both the villain and the hero. As the *accident prone driver* humans are fallible. Our current mental state will affect our approach to the driving task and to those around us, our age will affect response times, and we might use other equipment (e.g. a mobile phone) or take

illicit substances that reduce our performance further. Although the notion of the *accident prone driver* is almost consigned to history, the presumption that humans make mistakes is undoubtably a motivator for the introduction of greater automation. Human factors practitioners have raised concerns regarding the impact of automation on the human operator; with Bainbridge's "Ironies of automation" being seminal in this area [17]. In an automotive context, significant funding has been made available for human factors and driving research. As a consequence a substantial body of research regarding driver behaviour in relation to automation now exists. That said, accident statistics remain unacceptably high and media reports of fatalities involving partial automation are still depressingly frequent.

This view of the fallible driver contrasts with the automotive functional safety paradigm, which like road safety research has evolved slowly since the first E/E systems were fitted to vehicles in the 1980's. In this paradigm the driver is an integral part of vehicle control. Being integral to vehicle control and situationally aware, should a hazardous vehicle E/E system fault occur, the driver will typically mitigate the hazard and prevent an accident occurring. With the notion being that unless the hazard occurs in an operational situation where controllability is *difficult* or *uncontrollable*, greater than 90% of drivers will successfully retain vehicle control [162, Part 3 page 26]. The driver is clearly not fallible in this model.

This presents us with a dichotomy! The driver as the *problem* that the introduction of greater automation is seeking to solve, and the driver as the final *risk reduction measure*, who is typically capable of preventing vehicle hazards becoming accidents.

13.2 THE DRIVER AS A RISK REDUCTION MEASURE

13.2.1 *Automotive functional safety*

With its focus being to ensure that faults occurring within new E/E vehicle features fitted to a production vehicle do not adversely affect safety, the functional safety lifecycle process described by ISO 26262 focuses on the safety of individual vehicle features, each of which ISO 26262 defines as an *Item* [162].

Once a definition of the *Item* has been documented the functional safety lifecycle process typically starts with a hazard analysis. Being focused on the safety of individual vehicle features, an automotive hazard analysis will typically start with a list of system functions and foreseen operational situations. From this hazards will be identified and the risk of those hazards classified. As discussed previously, this is an effective process when the vehicle feature is self-contained and the driver remains within the vehicle control loop [109]. However, with the driver relinquishing aspects of the DDT, and safe vehicle feature behaviour potentially being influenced by external stimuli, the existing hazard analysis process is challenged [110].

This research hypothesises that this challenge is due in part to abstraction. Specifically that a hazard analysis process that begins by modelling a system as a set of functions, is attempting to model an AD system at a too lower level of abstraction. By taking a more systems engineering approach, advocates of methods such as STPA, suggest that the challenge of the analysis becoming too detailed too early is avoided. As an analysis method, STPA is certainly growing in popularity within the automotive industry. However, the complexity of automotive CSD observed together with the significant resource needed to construct such models, effective use of abstraction is not being utilised in practice. To support modelling the AD vehicle feature at a level of abstraction that facilitates vehicle level hazard analysis, this research asserts using the hierarchical control structure style version of the EVCM, together with the behavioural competency taxonomy. Used together in this way, the EVCM

provides the control hierarchy for the CSD, while the behavioural competency taxonomy provides the language from which to describe the behavioural competencies needed by the system (both human and automation) to operate the vehicle safely in its environment. Whilst the case studies conducted thus far suggest the EVCM as a viable alternative to creating a CSD from first principles, as discussed in Section 12.2, further work is needed to confirm that this approach does offer the envisaged positive benefits to the hazard analysis process.

13.2.2 *Maintaining safety*

The accident models underpinning automotive standards, such as ISO 26262 and ISO 21448, have a causal chain concept at their heart. With the notion that breaking the chain prevents a malfunction (i.e. a hardware, software or systematic fault in the context of functional safety, or a triggering condition in the context of SOTIF) leading to a hazard, and a hazardous event resulting in an accident.

The introduction of greater automation affects this relationship between a system malfunction, a hazardous event and ultimately an accident – particularly in relation to controllability. As the link between a hazard occurring and an accident taking place, controllability remains an important notion; both for functional safety and for SOTIF. The exception to this would be where vehicle control is completely delegated, as in the case of SAE Level 5 full automation. So it is reasonable to postulate that for SAE Levels 1 to 4 controllability remains relevant. However, of particular importance to AD safety is the fact that controllability, or to put it another way, the maintenance of vehicle safety, becomes a task that is shared between the driver and the automation. With the extent to which the task is shared, and the resultant responsibility split, being dependent on the level of automation deployed.

Although providing the benefits of a systems engineering view of AD vehicle feature accident causation, hazard analyses following the STPA method typically

models the human as the highest controller in the CSD hierarchy. A consequence of this is for the human to be modelled as a system *operator*, rather than as a part of the control system itself. This is a subtle but important distinction.

As discussed in Section 2.3.2 the use of automation in the automotive domain is atypical, because it is highly unlikely that once active the AD vehicle feature can be left unattended to safely complete its task. This research postulates that by placing the human as a controller at the top of the control hierarchy, as with *Classic* STPA, leads to the interaction between the human and automation to be modelled as an *aggregation* of task interactions. Although this may be appropriate for strategic level driving tasks, such as selecting a navigation route to follow or switching on a vehicle feature for the first time, this type of interaction is not appropriate for manoeuvring or control level tasks. For manoeuvring level tasks, such as *maintaining position in lane* or *changing lane*, the shared control required will be what philosophy research refers to as *shared agency*. For these types of task both agents need to work together on an aligned goal of maintaining vehicle safety, rather than their interaction simply being the aggregation of tasks.

FUTURE RESEARCH IDEA: With the EVCM and the *Shared Control* STPA method this research provides a conceptual model of *shared agency* in the context of AD systems. With the addition of the safety case argument pattern, this provides the practitioner with a method from which to reason about, and potentially to mitigate, hazards caused by shared control – and importantly, to articulate the level of confidence achieved. Identifying that for the task of driving many interactions between the driver and the automation involve shared agency raises further research questions. For example, what more can the automotive industry learn from shared agency and meshing sub-plans philosophy research that can be applied to AD shared control? As well as informing hazard analysis activities, could shared agency also inform AD design decisions or the regulatory framework for AD shared control? Or could shared agency and human-machine teaming research help engineering

teams better understand and design the relationship between the driver and the automation?

13.2.3 *The language of driving*

Seeking to further develop the MISRA VCM, to make a conceptual model applicable for AD features having shared control, highlighted the importance of explicit definitions. With the driving task effectively being split between the human driver and the automation, it was no longer appropriate to leave the interpretation of individual diagram elements to the reader. Particularly with regard to the important safety questions of: what functions or tasks are happening inside each element, who (i.e. which agent) is responsible for undertaking those tasks, who is responsible for monitoring the correct execution of those tasks, and ultimately who is responsible for maintaining the safety of the vehicle while those tasks are going on?

In an attempt to better define *driver control* (see Figure 1 page 19) a literature survey of driver models highlights a void in the literature. With the exception of Michon's HCM, few models were found that conceptualise the driver or the task of driving. Instead, driver modelling research appears to have focused on modelling a typical driver's performance for use in a vehicle simulation context. Based on the HTAoD, the behavioural competency taxonomy provides the analyst with a language to describe and reason about the task of driving at a level of abstraction that is familiar to many highly automated driving teams.

This research does not wish to claim that the behavioural competency taxonomy (see Appendix A page 210) is complete. Although, in each case study undertaken thus far the candidate system's behaviour has been successfully described using the current taxonomy, further case studies (see Section 12.1) would be required before a completeness claim could be made. However, the research has shown that further behavioural competencies can be added successfully should the need arise. Like other researchers [20, 173] this research

has focused on those behavioural competencies needed to complete the DDT and for the driver to interact with the vehicle's automation. What has not been considered is the behavioural competencies that the driver might exhibit when interacting with other road users, or that other road users might be exhibiting.

FUTURE RESEARCH IDEA: To be considered complete, the language of driving and shared control needs to extend beyond the boundary of an individual vehicle. If the behavioural competency taxonomy were extended to consider the actions of other road users, could shared control be explored between the driver and a pedestrian, or between the pedestrian and the automation? Then if an extended behavioural competency taxonomy were used in conjunction with multiple EVCMs, could the interaction between multiple road users be conceptualised? Again in this context, the shared agency could be occurring between machine agents (e.g. in the case of vehicle platooning), or between human and machine agents (e.g. a human driver following another vehicle that has a partial automation vehicle feature actively controlling that vehicle), or any number of other permutations between human and machine agents?

13.2.4 *Perception and timing*

A significant body of research exists discussing the potential impact of greater automation on the driver. For example, changing driver reaction times, changing skill levels, changing risk acceptance, etc. Hollnagel's Contextual Control Model (COCOM) provides a conceptualisation of the human decision making process, which allows the duration of event interpretation, action selection and action completion to be modelled and reasoned about. The *Shared Control* STPA method presented incorporates COCOM inspired questions into the loss scenario questions list (see Figure 17 page Figure 17), with the expectation being that by addressing these questions the analyst will identify loss scenarios caused by either event interpretation, action selection or action

execution delays. However, further case studies are required to establish whether the COCOM derived questions are a beneficial addition to the *Shared Control* STPA method.

FUTURE RESEARCH IDEA: One early research activity undertaken by the author involved creating a COCOM Matlab model, the aim of which was to explore the potential impact of timing in relation to the EVCM. Unfortunately, Matlab in the author's hands did not prove fruitful on this occasion. That set-back aside, the belief remains that there are timing aspects of shared control worth exploring. For example, in situations where the driver's perception of the environment will be very different to that of the automation's; occurring because each agent derives their information from completely different sources. As well as uncovering safety implications, could perception difference conceptualisation help the driver understand and interpret the automation's behaviour? From the perspective of safety, could this knowledge provide the driver with important insight about when to intervene because the automation's performance has deteriorated?

FUTURE RESEARCH IDEA: In addition to informing Human Machine Interface (HMI) design, could the EVCM and the behavioural competency taxonomy be used to explore other aspects of timing in relation to shared control? Section 5.2.1.2 introduced FRAM as a method that can be used to explore a system's resilience. Hollnagel suggests that socio-technical systems are intractable, meaning that often "the system in focus" cannot be fully described. For such intractable systems, Hollnagel suggests that it is "necessary to look for methods and approaches that can be used for systems that are incompletely described or underspecified" [80]. Consequently, could FRAM and the EVCM be used as the basis from which to reason about the resilience of a particular AD feature's shared control? Or to explore how the level of automation deployed (e.g. partial, conditional, etc.) affects the timing of HMI in the context of the EVCM?

FUTURE RESEARCH IDEA: Numerous research inquiries have addressed the removal of the driver from the vehicle control loop, and how that affects driver Situation Awareness (SA) and hence driver reaction time. In an AD driving context, Level 2 SA (i.e. comprehension of the current situation) and Level 3 SA (i.e. the projection of future states) are important facets of situation awareness. However, how does the level of automation deployed affect the timing in relation to COCOM, and can an analysis method be developed that helps safety practitioners reason about these effects?

It is clear that when Hollnagel created COCOM his focus was the human. However, in a machine learning context could the EVCM, the *Shared Control* STPA method and COCOM principles be used to reason about the machine's decision making process? For example, in the case of the well publicised and investigated Uber accident [117]? Although, the majority of accident blame was apportioned to the safety driver's lack of attentiveness, the Uber Advanced Technology Group not understanding the automated driving system's limitations was also said to be a contributing factor [117, Page v]. Therefore, could the EVCM and *Shared Control* STPA method be used in the context of machine learning, to explore the loss scenarios where the time to correctly classify objects could impede safety? In the case of the Uber accident, the fact that Elaine Herzberg crossed the road away from a designated crossing, while pushing her bicycle, meant the automated system was slow to correctly classify her as a pedestrian crossing in the path of the vehicle.

Moving the human from being the driver inside the vehicle, to being a Remote Operator external to the vehicle raises further perception questions. The author is aware of SAE Level 4 automation development projects where the ambition is to move the location of the safety driver from inside the vehicle to a remote location. Initially the Remote Operator would remain in vehicle line-of-sight, but with the ultimate aim being to locate them in a remote control room. For these advanced highly automated vehicle systems, the role of the Remote Operator will be to oversee the vehicle in situations where the automation has reached an impasse. For example, the vehicle's path is obstructed by a stationary object (e.g. a broken down vehicle). In such a situation the

Remote Operator would typically override the system, taking manual control to manoeuvre the vehicle around the obstacle. Once the obstacle has been successfully navigated vehicle control would be returned to the system.

FUTURE RESEARCH IDEA: These Remote Operator use cases raise safety questions regarding the Remote Operator's perception and what could affect their ability to perceive the vehicle's operating situation correctly. While the human remains seated in the vehicle their understanding of the vehicle's environment should be as good as that of the driver – providing they remain attentive of course! However, once outside of the vehicle their perception of the environment will change. When the Remote Operator moves outside of the vehicle will their relative position to the vehicle (i.e. the angle from which they are viewing the vehicle) affect their depth perception? Or when the Remote Operator is sitting in a remote control room, what additional information will they require to enable them to make correct decisions about the vehicle? The hypothesis here is that such errors in human perception are potential causes of hazards for a highly automated vehicle system. Although this is not strictly an example of shared control, could the EVCM and *Shared Control* STPA method be used together with inference hierarchy research (see Section 2.3.4.2 page 40) to explore the loss scenarios that could exist should the human Remote Operator make a wrong inference? Additionally, could such focused analyses identify human operator perception deficiencies that would need to be addressed for the human operator to make well informed vehicle control decisions?

13.3 DESIGNING FOR SHARED CONTROL

13.3.1 *Setting targets*

Having identified hazards and classified their risk, the next step in the automotive functional safety lifecycle involves defining safety goals. In an ISO 26262 context safety goals are the highest level safety requirements for

the *Item*. If the system implements the behaviour each safety goal describes, at the level of design integrity commensurate with the hazard risk, then the hazard risk will have been successfully reduced to an *acceptable level*.

In a manual driving context hazards are typically defined in relation to an unwanted vehicle property (e.g. unintended acceleration). Similarly, safety goals typically define the absence of the same 'unwanted' vehicle property phrased in relation to driver demand. So, an engine management system safety goal might be phrased as "the engine management system shall prevent vehicle unintended acceleration exceeding driver demand by $x ms^{-2}$ for longer than $y ms$". However, as highlighted during the Oxbotica workshop (see Appendix E), defining a hazard as *unintended acceleration* and framing the system's safety targets in relation to *driver demand*, is not appropriate in an AD system context. Instead, hazard definitions that describe vehicle behaviour that is observable from outside the vehicle are needed. For example, "The vehicle does not maintain a safe headway to preceding / adjacent vehicles". Similarly, in an AD context an STPA style safety constraint that "whilst in motion the subject vehicle shall maintain a safe headway to the preceding vehicle of $> x s$" is both observable and quantifiable. This research has presented a proof-of-concept method that promotes reasoning about the driving task at a level of abstraction commensurate with the behavioural competency taxonomy, rather than system level functions, which assists in identifying hazards at an appropriate level of abstraction for an AD system. In addition, observations made during the Oxbotica workshop suggest that the EVCM and the behavioural competency taxonomy provides a conceptual system model at an abstraction level that "makes sense to highly automated driving features that incorporate machine learning perception systems".

13.3.2 *Automotive risk management*

The automotive industry's approach to risk management has remained largely unchanged over decades. As discussed in Section 1.2.2, the automotive functional safety lifecycle focuses on the malfunctioning behaviour of individual

Items. With the implicit assumption being that individual *Items*, engineered to ISO 26262, will when integrated together create a vehicle that is safe. ISO 26262 defines *safety* as "the absence of unreasonable risk", and *unreasonable risk* as risk that is "judged to be unacceptable in a certain context according to valid societal moral concepts" [162]. Although an unquantifiable and arguably an impossible target to engineer to (let alone defend in a litigious marketplace), the automotive risk management uses residual risk as its measure. That is, if with safety measures in place the hazard risk can be classified as Quality Management (QM), then the risk that remains is deemed *acceptable*. Given that it will likely be infeasible to manipulate the severity of an accident, or the exposure to a particular operating situation, safety measures will typically modify controllability. As discussed in Section 6.2.1, maintaining controllability might be achieved by reducing engine power if a hazardous fault occurs within an engine control system, or limiting vehicle speed if because of an air suspension fault, the vehicle's ride height becomes stuck at the off-road height.

FUTURE RESEARCH IDEA: Using QM as an *acceptable* risk yardstick works while the driver remains within the control loop. However, in an AD context, what should the yardstick be, and how do automotive engineers determine that safety has been achieved? Given that AD systems incorporate perception systems, basing a safety case argument solely on risk feels both inadequate and inappropriate. A safety case argument for an AD vehicle feature must address the performance achieved by the perception system, which must introduce *uncertainty* into the argument. Currently, neither ISO 26262 nor ISO 21448 have any notion of uncertainty. This leaves a void in an automotive safety case argument for AD vehicle features.

Yes, as this research demonstrates, it is possible to conceptualise and analyse the nature of shared control in an AD context. By identifying loss scenarios and implementing functional modifications that prevent potentially hazardous shared control errors, a safety case argument for an AD vehicle feature's shared control can be made. However, this is but a small fragment of the safety case

argument needed for an AD vehicle feature. For an AD vehicle feature whose perception systems have a level of performance and undoubtable uncertainty, how should *unacceptable* be judged against *valid societal moral concepts*?

Understanding what will be acceptable when the driver is no longer responsible for vehicle control is a non-trivial endeavour. As discussed in Section 4.3, research suggests that the public's perception of risk is largely intuitive and rather than considering risk rationally, humans tend to use a combination of fact and feeling. Additionally, perception gaps clearly exist. With the public's perception of risk for the mundane task of driving undoubtably being at odds with the real risk. Also, with research suggesting that risk acceptance is less for voluntary rather than involuntary tasks, it is also fair to assume that the public will be even less inclined to accept risk when they have delegated the responsibility of driving to the automation. A premise for the introduction of greater automation is that when we as humans get behind the wheel we are fallible – we make mistakes. Uncertainty in the driving environment makes it improbable to conceive a time when an AD vehicle feature's perception system will be infallible. So how many AD vehicle feature perception errors will be judged as unacceptable in a certain context according to valid societal moral concepts? This is an open question not addressed by existing standards.

While the risks associated with adding new complex AD systems might be a risk management challenge, what cannot and should not be ignored is the potential benefits that greater automation can bring to personal transport. As discussed previously, the introduction of greater automation has benefits. From potentially reducing accidents caused by human error, through to supporting the mobility of an ageing population, and generally increasing mobility efficiency. So, is there a means by which the potential benefits of a new AD vehicle feature can be evaluated against the risks associated with deploying the technology – particularly for AD vehicle features that devolve the driver of vehicle safety responsibility for significant periods of the journey?

As discussed in Section 5.3.2 knowing when to stop and deciding that sufficient safety has been achieved is not purely an ALARP argument. In reality, ALARP is a difficult principle to argue and with it taking no account of the societal view

is open to criticism. In principle other factors such as corporate responsibility, ethical reasoning, potential business benefits and impacts will all affect the decision making process [155]. Although, perhaps equally as difficult to achieve, transparent communications are needed regarding the risks and benefits associated with the deployment of an AD vehicle feature – both from a regulatory perspective, but also from the perspective for the general public buying, using and interacting with such systems. Given these challenges, could the EVCM and the behavioural competency taxonomy be used to conceptualise and communicate risks and benefits transparently across multiple stakeholder groups?

13.3.3 *Scaling up*

ISO 26262 focuses on achieving the functional safety of an individual *Item*. With the implicit assumption being that by integrating of a number of functionally safe Items also yields a functionally safe vehicle. Although developed to address AD, the SOTIF process, described by ISO 21448, still focuses on individual vehicle features. But, as discussed throughout this thesis, a vehicle fitted with AD technology cannot be thought of in isolation. Therefore, work is needed to expand and *scale-up* the EVCM, behavioural competency taxonomy and *Shared Control* STPA method to consider multiple vehicles.

FUTURE RESEARCH IDEA: As touched upon in Section 13.2.3, the analysis of shared control and shared agency should expand beyond an individual vehicle boundary to consider agents located outside of the subject vehicle – both human and other machine agents. In addition, vehicle production systems are beginning to emerge that display information about the status of other vehicles, received via vehicle to vehicle communications. For example, the Polestar Connected Safety and Slippery Road Alert features provide the driver with information about crashes, breakdowns, and icy road conditions, all based on information received from other Polestar and Volvo vehicles on the road ahead [129]. One can envisage a time in the near future when AD vehicle features

will base their decision making on information received via vehicle to vehicle communications. This will necessitate the expansion of the hazard analysis process to include multiple vehicles. Therefore, future research is needed to determine how the EVCM could be used to conceptualise such a multiple vehicle system – perhaps by linking multiple EVCMs together?

14

CONCLUSION

The motivation for this research was the realisation that the introduction of greater vehicle automation would change the driving task and the driver's role. Changing driving both from the perspective of how the driver controls and interacts with the vehicle, but also from the perspective of what should be assumed about the driver's or automation's capability at any given time. Although greater vehicle automation is seen as the means by which *human error* is finally removed as a road traffic accident cause, the role of the human driver as the final safety mechanism in a complex vehicle control system cannot be understated. A career long interest in the notion of controllability, and the realisation that controllability and indeed functional safety, might need to be approached differently with greater automation, provided the final push needed to begin this research journey.

This thesis poses the research question: "how can the safety of AD be assured under different levels of shared human-vehicle control?" What started as a broad exploration of controllability, quickly became a focused consideration of shared control; specifically human and automation agent responsibility for a given driving task. Based on the MISRA VCM, the EVCM provides a conceptual model from which to reason about shared control in the context of AD. Key to the development of the EVCM was adding the hierarchy of control, needed to conceptualise the task of driving. To fully represent human cognition the EVCM includes perception and error blocks, with the error block being added to represent the need for automatic subconscious tasks to be brought into consciousness should something untoward happen.

Attempting to describe the activities underway within each EVCM block and the meaning of each signal flow arrow, highlights the ambiguity present in the original MISRA VCM – something that only becomes problematic once the

human driver no longer has sole responsibility for driving. Once the driving task is shared, the activities being undertaken need stating explicitly, as does the responsibility for undertaking each activity. The behavioural competency taxonomy provides the vocabulary from which to describe the complete driving task – both the near instantaneous control level tasks, but also the longer duration manoeuvring and strategic level tasks. This ensures that adequate consideration is given to all tasks, previously undertaken solely by the human driver, and needed to safely control a vehicle in the environment, and not just the DDT related tasks as is typically the case.

To address the research question three research products have been created: a vehicle model and behavioural competency taxonomy that allows AD shared control to be conceptualised, a hazard analysis method for analysing shared control hazard causes, and an accompanying safety case argument pattern. To evaluate the *effectiveness* of the *Shared Control* STPA method a set of case studies has been undertaken to establish the extent to which the research products exhibit the properties summarised as:

- Used together the EVCM and the behavioural competency taxonomy supports AD hazard analysis by creating a conceptual model of an AD vehicle feature's shared control.
- This conceptual model then forms the CSD for the *Shared Control* STPA method, which being based on cognitive principles supports the identification of hazard causes, which might otherwise have remained undiscovered.
- Finally, the supporting safety case argument method highlights potential loss scenarios attributable to shared control and emphasises how AD feature design modifications mitigate such hazard causes.

The case studies have shown how AD vehicle features, having differing levels of automation, can be modelled using the EVCM and behavioural competency taxonomy. The assertion is that the taxonomy is complete for a single vehicle, but further evaluation work would be needed to confirm this to be the case. The need to expand the behavioural competency taxonomy beyond an individual

vehicle has also been acknowledged (see Section 12.5) and is suggested as potential future research.

The evaluation of the *Shared Control* STPA method is less conclusive. The case studies undertaken show the *Shared Control* STPA method to be a proof-of-concept. However, these cases are not expansive enough to discern whether the method can help the analyst identify hazard causes that would have remained undiscovered had a different hazard analysis method been used.

It is acknowledged that much of the evaluation has been undertaken solely by the author, which does raise questions about the utility of the method when used by others. The case studies have shown the method's utility in uncovering new potential errors in AD vehicle feature shared control that could be hazardous. However, it is impossible to determine the extent to which this positive outcome is the result of a 25 year automotive career, rather than the *effectiveness* of the EVCM and the *Shared Control* STPA method. That said, feedback from the engineering team at Oxbotica is positive. At the time of writing, the Oxbotica team continues to use the EVCM, behavioural competency taxonomy and *Shared Control* STPA method to explore the interaction between the Remote Operator and the Selenium system's automation.

The need to scale the automotive hazard analysis process beyond a single vehicle was understood from the beginning (see Chapter 3). However, before undertaking hazard analysis at a road network level of abstraction (i.e. vehicles interacting with other vehicles and other road users) an efficient way of analysing hazards associated with one vehicle was needed. This thesis contributes a proof-of-concept method for analysing hazard causes in an individual vehicle whose safety relies on shared control. Further developments that could be made to this research contribution to answer further research questions has been discussed, together with the potential to scale-up the EVCM to conceptualise and analyse shared agency in a multi road users driving environment. The author continues to research in this area and invites others to contribute to the interesting question of shared agency in a highly automated driving context.

The EVCM could also have utility beyond the automotive domain. This research has focused solely on the vehicle as the system under control. However, replacing the vehicle plant model with a plant model from another domain is equally conceivable. In addition, there is no reason to assume that the replacement plant model would have to be another transport system. For example, during a discussion with colleagues at the University of York the potential for the EVCM to conceptualise potential hazards associated with administering drugs within a medical Intensive Care Unit context was hypothesised.

Controllability remains an important consideration for AD vehicle feature safety. Particularly the composition of shared control, while the human driver retains responsibility for aspects of vehicle control and safety. This research has not only highlighted the importance of being explicit about the responsibility for shared control, but has also highlighted to the author that subtleties exist regarding the nature of shared control. Contemporary hazard analysis methods, like STPA, typically model the human-machine interaction as an aggregation of tasks. However, the human-machine interaction that occurs between the driver and vehicle automation is a *deliberate act* between two collaborating actors – what philosophers refer to as *shared agency*. The notion of driving shared control as *shared agency* is the topic of ongoing research by the author and colleagues at the University of York.

This research has sought to answer the question "how can the safety of AD be assured under different levels of shared human-vehicle control?" and in doing so makes the following contribution to the state of human knowledge:

1. Reviewing vehicle safety, risk management and controllability literature, from the perspective of an automotive functional safety practitioner seeking to achieve AD safety assurance, has highlighted functional safety lifecycle tools, techniques and assumptions requiring redress.

2. The development of a vehicle model and behavioural competency taxonomy that aids the automotive functional safety practitioner to conceptualise and describe the nature of an AD vehicle feature's shared control.

3. The development of a concrete hazard analysis method that uses the vehicle model and behavioural competency taxonomy to analyse shared control hazard causes, together with the creation of a safety case argument pattern for that.

4. A series of case studies has established that the method presented exhibits the qualities necessary to be deemed a proof-of-concept. However, further evaluation work is needed to establish the method's viability as a hazard analysis method for automotive functional safety practitioners working in a product development environment.

The author would now like to invite others to take the EVCM and its associated ideas and to experiment further. Perhaps to explore shared interactions between vehicles, or between a highly automated vehicle and more vulnerable road users like cyclists and pedestrians. Or maybe to even transplant the EVCM's ideas and concepts into other domains, such as maritime or medical.

Part VI

SUPPORTING MATERIAL

A

BEHAVIOURAL COMPETENCIES

In a highly automated driving context this controlling functionality is typically referred to as the AD feature's *'behavioural competencies'*. It is understood that the SAE is currently working to harmonise behavioural competencies [173] for highly automated driving, that will likely be derived from contemporary verification and validation focused research [119, 189]. Having reviewed this research, together with the earlier HTAoD taxonomy [186] and insight gained from the evaluation case studies, a list of generic driving behavioural competency has been developed for the EVCM.

The below behavioural competencies are grouped following Michon's hierarchy of *strategic, manoeuvring* and *control* levels. With the main distinction between these three categories being the time available to the agent to interpret, decide and to take action. Typically *strategic* tasks are afforded several seconds to complete, *manoeuvring* tasks one or two seconds, while *control* level tasks are typically performed in milliseconds.

Researchers typically split (e.g. [173, 186]) the *manoeuvring* level into two further categories: *tactical* and *operational*. Here, *operational* behaviours are those predominantly 'rule based' DDT activities undertaken to control the vehicle and to achieve the mission's *strategic* goals. For example, *maintain speed, maintain position in lane* or *navigate junctions*. While *tactical* tasks are actions taken in response to something happening and so often involve interacting with another road user. For example, *right-of-way decision, deal with different road types* and *emergency manoeuvres*. In either case, the Sense Understand Decide Act (SUDA) decision making process will typically only last a few seconds.

A.1 STRATEGIC LEVEL COMPETENCIES

ROUTE PLANNING Often referred to simply as 'navigation'. This competency involves determining the route to be followed. This might involve using prior knowledge or available resources such as map or traffic data. This task is typically done before the journey starts, but may be performed again mid-route if external factors dictate a change – for example, an unexpected road closure.

PERFORM SURVEILLANCE Is able to sense the environment in which the vehicle operates and also the vehicle's behaviour within that environment. Understands how artefacts (both static and dynamic) in the vehicle's environment might affect its safe operation or progress towards the destination. Decides what changes (if any) are needed to vehicle operation to maintain progress and minimise risk. For example, reducing speed, increasing braking distances and electing not to use automation as traffic density increases, or increasing braking distance to the vehicle ahead when an aggressively driven vehicle is close behind.

COMPLY WITH RULES Acts on advice / instructions / rules / guidance provided in relation to road traffic laws (e.g. the Highway Code [39]). Is able to respond to directions / instructions from authorised personnel (e.g. police officer, pedestrian crossing guard "Lollipop' man / lady"). Uses this information in the context of traffic and environmental conditions to select a desirable route, assess progress towards destination, etc.

RESPOND TO TRAFFIC CONDITIONS Is able to sense the vehicle's operating environment (e.g. weather, visibility, sun glare) and understands how this could influence vehicle performance and capability. Decides what changes (if any) need to be made to the vehicle operation to maintain safety and progress towards the destination. For example, reducing speed and increasing braking distances in cold weather or poor visibility.

USE AUTOMATION APPROPRIATELY Typified as a human behavioural competency today, one could imagine a future time when responsibility for this competency could be shared. The competency requires a comprehension of both the vehicle's operating environment and the capability of the automation. But importantly, it also requires an understanding of how those external environmental factors might affect the automation's performance. Thus, vehicle safety is assured by only using automation to achieve vehicle control when it is known that the automation will achieve this aim successfully.

A.2 MANOEUVRING LEVEL COMPETENCIES

A.2.1 *Tactical*

PULL AWAY FROM STANDSTILL Deals with launching the vehicle from a standstill. May include ascertaining whether the vehicle's path is clear, coordinating the release of the vehicle's parking brake, preventing the vehicle from rolling backwards, and coordinating the release of the braking torque and the application of propulsion torque.

RIGHT-OF-WAY DECISION Is able to sense traffic control devices, signage and infrastructure details. Understands how to interpret the data received from these traffic artefacts to determine the safe reference trajectory and speed for the vehicle. Decides what changes (if any) are required to the vehicle's reference trajectory and target speed to maintain safety. In addition, is able to sense the position of other road users in the vehicle's environment and based on their categorisation understands how to interact with them – including any relevant prioritisation.

DEAL WITH DIFFERENT ROAD TYPES Is able to sense traffic control devices, and interpret signage and infrastructure details. Decides what changes (if any) are required to the vehicle's trajectory and target speed to maintain safety.

AVOID OBSTACLES Is able to sense objects in the vehicle's path (both static and dynamic) and understands which of the objects identified pose a collision risk. Decides what lateral adjustment and longitudinal acceleration rate changes (typically negative acceleration) are required to avoid those obstacles deemed a collision risk.

EMERGENCY MANOEUVRE Is able to sense the environment around the vehicle and determine when an unsafe situation exists. In this context an unsafe situation could be an imminent collision risk, but it could also be the subject vehicle losing control (e.g. hitting a patch of black ice). This behavioural competency must decide the safe exit path for the subject vehicle, modifying the vehicle's speed and direction accordingly, to maintain safety.

NAVIGATE TEMPORARY CONDITIONS This competency deals with the atypical temporary conditions which may be encountered. This includes the ability to detect and correctly respond to work zones (e.g. road works) and to changes resulting from unplanned events (e.g. a road traffic accident). This behavioural competency can detect and correctly respond to temporary signage and direction markers (e.g. temporary speed limit, detours, traffic cones), as well as from people directing traffic (e.g construction zone workers, police officers, first responders).

BRING VEHICLE TO A STOP This behavioural competency deals with bringing the vehicle to a controlled stop, and may include holding the vehicle stationary (i.e. with the parking brake) once the vehicle speed has reached zero. This behavioural competency may also include monitoring the speed and distance of succeeding vehicles. In doing so, the vehicle's deceleration rate can be managed to not only maintain sufficient headway to the preceding vehicle, but also to ensure that the succeeding vehicle is afforded sufficient time to brake safely.

PARK THE VEHICLE As the name suggests, this competency deals with parking the vehicle in a designated parking bay, using one of the standard parking manoeuvres: parallel parking, reversing into a bay, or driving forwards into a bay. May include identifying the location of a parking space, and determining the suitability of the space found – e.g. is the space big enough to accommodate the subject vehicle.

A.2.2 *Operational*

MAINTAIN SPEED Is able to sense the vehicle's current speed and understands the longitudinal acceleration rate (*Note:* negative rate implies a deceleration) needed to maintain the current desired vehicle speed. Decides the longitudinal acceleration rate needed to maintain the vehicle at its desired speed.

FOLLOW OTHER VEHICLES Is able to sense a preceding vehicle and understands the relative speed and distance to the preceding vehicle. Decides what changes are required to the vehicle longitudinal acceleration (*Note:* negative rate implies a deceleration rate) to maintain a consistent gap to the preceding vehicle.

MAINTAIN POSITION IN LANE Is able to sense the lane boundaries and the vehicle's relative position in lane. Understands the vehicle's optimal position in lane (reference trajectory) to maintain safety (e.g. distance to static / dynamic objects, future manoeuvre). Decides what lateral adjustment is needed to follow the reference trajectory.

OVERTAKE / CHANGE LANE When there is a requirement to change lanes (typically when approaching and wishing to pass a slower vehicle), is able to determine when the adjacent lane is clear and so is safe to manoeuvre into. This may include managing the subject vehicle's speed from behind the slower

vehicle, into the potentially faster adjacent lane. The competency includes the ability to safely move into left or right hand adjacent lanes.

ENHANCE VEHICLE CONSPICUOUSNESS Deals with making the vehicle easily seen by other road users. For example, the use of front and rear side-lights. Also included in this competency is making the subject vehicle's intentions understood by other road users. For example, the correct use of brake lights and directional indicators.

NAVIGATE JUNCTIONS / ROUNDABOUTS This competency deals with the detection and correct negotiation of road junctions. This includes the correct approach to the junction, identifying the type of junction (e.g. T-junction, cross-roads, roundabout), an understanding of the correct interaction with other road users at the junction (e.g. priorities) and the rules for safely navigating the subject vehicle through the junction.

PERFORM MERGE / NAVIGATE ON / OFF RAMPS This competency deals with matching the subject vehicle's speed to allow a safe merge to happen. This could either be when the subject vehicle is entering or exiting a multi-lane highway and is required to either merge into, or merge out of the traffic flow. Or when the vehicle in an adjacent lane is indicating its intention to change lanes and the subject vehicle modifies its speed to allow the merge to take place safely.

NAVIGATE CROSSINGS: PEDESTRIAN / RAIL Is able to detect the presence of the crossing and safely negotiate the subject vehicle through the crossing. The types of pedestrian crossing dealt with will depend on the region of operation and on the ODD. For pedestrian crossings this could include fully manual 'zebra' crossings or crossings under traffic light control (e.g. pelican, toucan) [169]. A range of railway crossing types is possible and the types encountered will again be dependent on the region of operation and the ODD.

The competency includes the rules for safely negotiating railway crossing – such as, not entering the crossing until the exit is clear.

PERFORM U-TURN / N-POINT TURN A behavioural competency applicable to urban settings. The n-point turn refers to the ability to turn the vehicle around in the road, using forward and reverse gears, until it is facing in the opposite direction. Typically, a manoeuvre reserved for very quiet streets. The u-turn is specific to roads with a central reservation, where the subject vehicle drives through a gap in the central reservation to continue its journey in the opposite direction. Depending on the location and region, this manoeuvre might be under traffic light control, or it might require the subject vehicle to determine when it is safe to pull from the central reservation into the live lane. In either case, the competency will include identifying where and when it is both legal and feasible to perform the manoeuvre.

TRANSITION OF CONTROL This behavioural competency deals with the successful transfer of DDT control between the human agent and the automation and is applicable to all automation levels below SAE Level 5. For a highly automated system (i.e. SAE Level 4) the transition of control might only take place once at the beginning of the mission. However, for lower levels of automation (i.e. ADAS, partial and conditional automation) this transfer will happen more frequently. Typically, a combination of audible, visual and haptic cues are used to communicate the status and progress of the control transition.

DRIVER ATTENTIVENESS Often used in conjunction with the above behavioural competency, the driver attentiveness behavioural competency deals with driver attentiveness and situational awareness determination. Information that is needed to ensure the human driver can successfully undertake the vehicle control task. A typical application of this behavioural competency is in the context of partial or conditional automation. For example, the functionality needed to ascertain that the driver has 'eyes on the road' before returning DDT

control to them. However, driver warning systems (e.g. suggesting the driver take a coffee break) and vehicle inhibit devices (e.g. a breath analyser that is required to give a negative alcohol or drugs result before the vehicle can be started) would also fall under this heading.

A.3 CONTROL LEVEL BEHAVIOURAL COMPETENCIES

HOLD VEHICLE STATIONARY Is able to prevent the vehicle from moving when required. This will typically include the correct application of the vehicle's foundation brakes to hold the vehicle still, either on a flat surface or on a gradient. Might also include the application of the vehicle's parking brake when required. Depending on the vehicle architecture and regulatory requirements, may include the functionality to safely apply the parking brake while the vehicle is in motion to maintain fall-back brake capability.

PERFORM LATERAL (STEERING) CONTROL Is able to sense the position of the vehicle's steering actuators and understands the transfer function needed to adjust them to achieve the required vehicle lateral motion. May also incorporate feedback, capable of sensing the vehicle's previous response to steering actuation and modifying the transfer function accordingly. Decides the new steering actuator position needed to achieve the required lateral motion.

PERFORM LONGITUDINAL (ACCEL.) CONTROL Is able to sense the current position of the vehicle's acceleration actuator input and understands the change in input needed (transfer function) to achieve the desired longitudinal acceleration. May also incorporate feedback, making it capable of understanding the vehicle's previous response to the actuation, and modifying the transfer function accordingly. Decides the new acceleration actuator input needed to achieve the required longitudinal acceleration.

PERFORM LONGITUDINAL (DECEL.) CONTROL Is able to sense the current position of the vehicle's deceleration actuator input and understands the change in input needed (transfer function) to achieve the desired longitudinal acceleration (*Note:* negative acceleration in this case). May also incorporate feedback, making it capable of understanding the vehicle's previous response to the actuation, and modifying the transfer function accordingly. Decides the new deceleration actuator input needed to achieve the required longitudinal acceleration.

REVERSE THE VEHICLE Deals with manoeuvring the vehicle in reverse gear. May simply involve applying propulsion torque to propel the vehicle in the reverse direction. However, could also include the detection of objects in the vehicle's path and rear cross traffic collision detection.

A.4 PRE AND POST DRIVING BEHAVIOURAL COMPETENCIES

A.4.1 *Pre-driving*

PERFORM PRE-OPERATIVE TASKS These are the checks that should be carried out, at least daily, before the vehicle is used. Although driving a vehicle that is not "roadworthy" is a punishable driving offence in many regions [39, check-vehicle-safe], these daily checks are often overlooked. Roadworthiness includes checking daily that the windscreen, windows and mirrors are clear, all lights work and the brakes work. Additionally, checking other vehicle systems in accordance to the vehicle's handbook – e.g. engine oil level, brake fluid, battery. In a manual driving context this behavioural competency also includes ensuring that all driver controls are correctly positioned to drive – i.e. mirror, steering wheel and seat positions.

START THE VEHICLE For some vehicles this behavioural competency may no longer be applicable. Involves placing the key into the ignition and starting

the vehicle's power train. Where applicable, the vehicle's dashboard lights should be checked to ensure all warning lights have distinguished before beginning to drive. The journey should be aborted and the vehicle rectified if warning lights fail to extinguish.

A.4.2 *Post-driving*

MAKE THE VEHICLE SAFE This behavioural competency relates to making the vehicle safe after it has been brought to a stop and parked. Depending on the vehicle type this might include: checking that the parking brake is applied, turning off electrical systems, checking that all windows and doors are closed, and that the vehicle is locked.

		Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety
Pre	**Perform pre-operative tasks**	Perform pre-operative tasks			Start the vehicle																	
Control	**Perform control tasks**	Hold the vehicle stationary			Perform lateral (steering) control			Perform longitudinal (accel.) control			Perform longitudinal (decel.) control			Reverse the vehicle								
Manoeuvring	**Perform operational manoeuvring tasks**	Maintain speed			Follow other vehicles			Maintain position in lane			Overtake other vehicles / change lane			Enhancing vehicle conspicuousness			Navigate junctions / roundabouts			Perform merge / navigate on_off ramps		
	Perform tactical manoeuvring tasks	Pull away from standstill			Right-of-way decisions			Deal with different road types			Avoid obstacles			Emergency manoeuvre			Navigate temporary conditions			Bring vehicle to a stop		
Strategic	**Perform strategic tasks**	Route planning			Perform surveillance			Comply with rules			Respond to traffic conditions			Use automation appropriately								
Post	**Perform post-operative tasks**	Make vehicle safe																				

		Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety	Does	Monitors	Achieves Safety
Pre	**Perform pre-operative tasks**																					
Control	**Perform control tasks**																			Perform u turn / n-point turn		
Manoeuvring	**Perform operational manoeuvring tasks**																Navigate crossing: pedestrian / rail					
	Perform tactical manoeuvring tasks																					
Strategic	**Perform strategic tasks**																			Transition of control		
Post	**Perform post-operative tasks**																					Driver awareness / attentiveness

Key:

Not in scope	
In scope: automation	A
In scope: human	H
In scope: shared	S

Figure 30: Driving behavioural competencies

B

VEHICLE FEATURE USE CASES

Based on manufacturers production or research vehicle features, this section describes a number of vehicle feature use cases. The features described have varying levels of automation, from SAE Level 1 through to SAE Level 4 (see Section 2.3.1, page 31). These use cases are used throughout the thesis to provide examples from which to illustrate research ideas and concepts.

B.1 ADAPTIVE CRUISE CONTROL (ACC)

B.1.1 *Overview*

ACC is designated as an SAE Level 1 system [141]. Like the cruise control feature, when active ACC takes control of vehicle speed from the driver. However, ACC also has the ability to detect the distance (headway) to the vehicle ahead. Therefore, ACC is able to reduce the vehicle speed (below the SET speed) to maintain a safe headway to the preceding vehicle. At any time the driver can override ACC with either the accelerator or brake pedals.

GOAL: To control vehicle speed and automatically maintain the distance to the vehicle in front.

BRIEF DESCRIPTION: When conditions allow, the Driver activates ACC for the first time by pressing the SET + button when the desired speed is reached. The Vehicle maintains that desired speed and headway to the vehicle in front; this may require the vehicle to apply the brakes and slow the vehicle or bring it to a complete stop. The Driver may use SET + and SET - to adjust the current vehicle speed, or may press the accelerator pedal

ACTOR	ROLE	INTEREST / CONCERN
Driver	Uses driver inputs to control the ACC system	Activates ACC when required via the SET + or RESUME buttons. Is able to cancel ACC by pressing either the CANCEL button, or the brake pedal. Incremental changes to vehicle speed can be made by pressing SET + or SET -. Headway to the vehicle in front may be increased or reduced. May momentarily override set speed by pressing the accelerator pedal while ACC is active.
Vehicle Environment	Dictates when it is safe for the driver to use ACC	Conditions may exist where it would not be appropriate for the driver to use ACC. For example, on a very windy road or in conditions like fog when vehicle sensors may fail to detect nearby objects.
Vehicle Manufacturer	Calibrates the ACC response	Requires vehicle behaviour, refinement and economy to be the same during ACC operation as normal driving. Requires the ACC system to behave consistently, thus promoting customer acceptance and trust.

Table 6: ACC actor and stakeholder interests

to momentarily override the current *Set Speed*. The driver presses the CANCEL button to return to manual control. Pressing the brake pedal at any time will cancel ACC. If RESUME is pressed during the drive cycle then ACC will reactivate at the last *Set Speed*. The Vehicle will illuminate the headway driver warning to indicate when a preceding vehicle has been detected. In certain circumstances the vehicle may require the Driver to resume manual control, perhaps when the preceding vehicle makes an emergency stop. In such cases an audible and visual warning will be given.

PRE-CONDITIONS: The Vehicle is being driven and the speed is being controlled by the Driver manually.

TRIGGER: The driver presses SET + or RESUME

B.1.2 *Requirements*

REQ 1: The ACC system conforms to the functional and performance requirements of ISO 22179:2009.

REQ 2: There shall be a minimum speed defined, below which ACC cannot be activated; thus discouraging ACC use at low speeds when vulnerable road users maybe in close proximity to the vehicle and where the Driver may have insufficient time / space to react if required to do so.

REQ 3: The CANCEL button or brake pedal inputs shall always take precedence over other ACC driver inputs, thus ensuring that the Driver is always able to cancel cruise if required.

REQ 4: An audible and visual *Driver Override* warning shall be used to indicate to the Driver when a transition back to manual control is required.

REQ 5: A visual *Headway* driver warning shall be used to indicate to the Driver when the system has detected a preceding vehicle and is operating in a headway controlling mode.

B.1.3 *Use case: main flow*

1. The Driver presses the SET + [*Alternate Flow 1:Press SET +*] or RESUME [*Alternate Flow 2:RESUME*] button to indicate they wish to activate ACC [*Exception Flow 3:Incorrect ACC Use*]
2. The Vehicle controls powertrain torque to maintain the *Set Speed*.
3. If the Vehicle RADAR detects a target ahead (preceding vehicle), illuminates the *Headway* driver warning and then controls powertrain torque and brake pressure to maintain the *Headway* to the target [*Alternate Flow 3:Target Disappears*][*Alternate Flow 4:Vehicle Brought to a Stop*][*Alternate Flow 4:Brake Pressure Limit Exceeded*] [*Exception Flow 1:Incorrect Target Discrimination*][*Exception Flow 2:Target Undetected*]
4. The Driver monitors the Vehicle Environment [*Exception Flow 3:Incorrect Cruise Use*] and may:
   a) change the *Set Speed* using the SET + [*Alternate Flow 1:Press SET +*] or SET - [*Alternate Flow 3:SET -*] buttons,
   b) momentarily override the current *Set Speed* using the accelerator pedal [*Alternate Flow 4:Accelerator Press*], or
   c) deactivates ACC by pressing the CANCEL button or brake pedal [*Alternate Flow 5: Cancel ACC*] as required.
5. The Vehicle remembers the last *Set Speed* for the remainder of the drive cycle.

POST CONDITIONS: ACC is active, the *Set Speed* is stored in memory and the Driver controls the Vehicle manually.

B.1.4 *Alternative flow 1: press SET +*

DESCRIPTION: Pressing SET + when ACC is inactive sets the *Set Speed* to the current vehicle speed and activates ACC. However, if SET + is pressed while ACC is active then *Set Speed* is incremented by a calibrated amount.

1. At Step 1:
    a) If ACC is inactive then the Vehicle activates ACC and sets *Set Speed* to the current vehicle speed.
    b) If ACC is active then the Vehicle increments the *Set Speed* by the predetermined amount
    c) Rejoins the main flow at step 2

B.1.5 *Alternate flow 2: press RESUME*

DESCRIPTION: Pressing RESUME reactivates ACC at the previous *Set Speed* stored in memory. If no *Set Speed* is stored then no action is taken.

1. At Step 1:
    a) If a *Set Speed* is stored the Vehicle activates ACC and rejoins main flow at step 2.
    b) Else ACC remains inactive and the use case ends.

POST CONDITIONS: ACC is inactive, no *Set Speed* is stored, and the Driver controls vehicle speed manually.

B.1.6 *Alternate flow 3: target disappears*

DESCRIPTION: The Vehicle may be controlling the *Headway* to the target ahead when that target disappears. Perhaps the preceding vehicle has turned

off the road or moved into another lane. When this situation occurs ACC must transition gracefully from headway control back into speed control mode; a graceful transition is required to promote Driver comfort.

1. At Step 3:

    a) The Vehicle can no longer detect the target ahead and extinguishes the *Headway* driver warning

    b) If ACC was using brake pressure to control the *Headway* then the brake pressure is released slowly.

    c) The vehicle increases powertrain torque until the vehicle speed reaches the *Set Speed*.

    d) The use case rejoins the Main Flow at step 2 (speed control mode).

B.1.7 *Alternate flow 4: vehicle brought to a stop*

DESCRIPTION: The target vehicle slows to a complete stop. The ACC system holds the Vehicle stationary until the Driver presses the accelerator pedal (see Note 3). At which point *Headway* control resumes.

1. At Step 3:

    a) The Vehicle comes to a complete stop and the brake pressure is maintained to hold the Vehicle stationary.

    b) The distance to the target grows as the preceding vehicle moves off.

    c) The Vehicle remains stationary until the Driver presses the accelerator pedal to indicate that it is safe for the vehicle to pull away.

    d) The use case rejoins the Main Flow at step 3 (headway control mode).

B.1.8 *Alternate flow 5: brake pressure limit exceeded*

DESCRIPTION: The *Maximum Braking Force* that ACC is allowed to apply automatically is limited. Thus if the preceding vehicle decelerates rapidly (e.g. making an emergency stop) then ACC will disengage; requiring the Driver to apply manual brake force in order to avoid a collision.

1. At Step 3:
    a) The Vehicle detects the rapid deceleration of the target vehicle ahead.
    b) Being unable to maintain the *Headway* to the target vehicle without exceeding the *Maximum Braking Force* the Vehicle indicates that *Driver Override* is required and disengages cruise.
    c) On receiving the *Driver Override* warning the Driver applies force to the brake pedal in order to avoid hitting the preceding vehicle that has just stopped abruptly.
    d) The use case ends.

POST CONDITIONS: ACC is inactive, the *Set Speed* is stored in memory, and the Driver controls vehicle speed manually.

B.1.9 *Exception flow 1: incorrect target discrimination*

DESCRIPTION: The Vehicle must correctly reject RADAR targets that are stationary objects (e.g. roadside furniture, parked cars, bridges, etc.) and vehicles travelling in other lanes. If a target is not correctly discriminated then the Vehicle may slow or stop when it is not required to do so. This may result in both Driver frustration, and confusion for other road users.

1. At Step 3:
    a) The Vehicle incorrectly interprets the RADAR reflection received as being a target vehicle travelling in the lane ahead.

b) The Vehicle illuminates the *Headway* driver warning to indicate to the Driver that a target vehicle has been detected.

c) The Vehicle transitions into headway control mode and begins to slow the vehicle (reduction in powertrain torque and application of brake pressure) to stop behind the supposed target vehicle.

d) The Driver realises that the Vehicle is incorrectly slowing or stopping and takes mitigating action:

   i. Driver overrides ACC by pressing the accelerator pedal – use case goes to [*Alternate Flow 7: Accelerator Press*]

   ii. Driver cancels ACC by pressing either the CANCEL button or the brake pedal – use case jumps to [*Alternate Flow 8: Cancel ACC*]

B.1.10 *Exception flow 2: target undetected*

DESCRIPTION: The system fails to detect the presence of a target ahead – perhaps environmental conditions (like fog) are adversely affecting the RADAR's sensing abilities. As a consequence ACC does not maintain the correct *headway* to the target vehicle ahead. Therefore, the Driver must manually intervene to avoid an accident.

1. At Step 3:

   a) The Vehicle fails to detect the presence of a target vehicle ahead and continues to operate in speed control mode.

   b) The Driver becomes aware that the Vehicle has not detected the target vehicle ahead (*Headway* driver warning does not illuminate or the headway to the vehicle ahead reducing) and cancels ACC to mitigate; either by pressing the CANCEL button or the brake pedal – use case jumps to [*Alternate Flow 8: Cancel ACC*]

B.1.11 *Alternate flow 6: SET -*

DESCRIPTION: If SET - is pressed while ACC is active then the *Set Speed* is decremented by a calibrated amount. If ACC is inactive then pressing the SET - has no effect.

1. At Step 3:
    a) If ACC is active then the Vehicle decrements the *Set Speed* by the predetermined amount and rejoins the main flow at step 2
    b) Else ACC remains inactive and the use case ends.

POST CONDITIONS: ACC is inactive, a *Set Speed* may be stored in memory, and the Driver controls vehicle speed manually.

B.1.12 *Alternate flow 7: accelerator press*

DESCRIPTION: The Driver wishes to override the *Set Speed* for a short time – perhaps to overtake the vehicle ahead.

1. At Step 3:
    a) The Driver presses the accelerator pedal to indicate their wish to increase vehicle speed above the *Set Speed*
    b) The Vehicle controls powertrain torque in response to the Driver's accelerator input
    c) The Driver releases the accelerator pedal and the use case rejoins at step 2.

B.1.13 *Alternate flow 8: cancel ACC*

DESCRIPTION: The Driver may regain manual vehicle speed control at any time by pressing either the CANCEL button or the brake pedal. In either case the *Set Speed* remains stored in memory for the remainder of the drive cycle.

1. At Step 2 or Step 3:
    a) The Driver presses either the CANCEL button or the brake pedal indicating they wish to cancel ACC
    b) The Vehicle deactivates ACC control and remembers the *Set Speed* for the remainder of the drive cycle
    c) The use case ends.

POST CONDITIONS: ACC control is inactive, a *Set Speed* is stored, and the Driver controls vehicle speed manually.

B.1.14 *Exception flow 3: incorrect ACC use*

DESCRIPTION: Safe vehicle operation requires the Driver to remain vigilant while cruise is active, modifying the *Set Speed* or cancelling ACC as the Vehicle Environment dictates. Failure to react to the changing Vehicle Environment may result in inconsistent system behaviour; potentially leading to Driver confusion, increased workload and maybe even an accident.

1. At Step 1 or Step 4:
    a) The Driver either elects to engage ACC when the Vehicle Environment is not appropriate (step 1), or fails to give the correct ACC control inputs needed (modify *Set Speed*, cancel) (step 4) to maintain safe vehicle operation

b) The Vehicle controls powertrain torque and brake pressure to maintain either the *Headway* to the target (if applicable) or the *Set Speed*.

c) The Vehicle makes targeting errors leading to inconsistent ACC behaviour. This could lead to the need for greater Driver intervention, which could potentially lead to accident.

POST CONDITIONS: Worse case a vehicle accident results.

B.2 HIGHWAY ASSIST SYSTEM (HAS)

B.2.1 *Overview*

HAS is designated as an SAE Level 2 system [141]. Like ACC it allows the Driver to set a desired vehicle speed which the automation will control too. As well as controlling vehicle speed to that set by the Driver, or to a speed that allows the vehicle to safely follow another vehicle, HAS also controls the lateral position of the vehicle. Normally the *Required Trajectory* determined by the system will be the lane centre. However, if the vehicle detects that the vehicle ahead has moved away from the lane centre (e.g. because they are overtaking a wide vehicle or vulnerable road user like a motorcyclist) then HAS will adjust the *Required Trajectory* accordingly.

The below use case has been written to reflect the behaviour of the Alfa Romeo Highway Assist System [6, 25].

GOAL: To control vehicle speed and direction, thus automatically maintaining the vehicle's safe position in lane and distance to the vehicle in front.

BRIEF DESCRIPTION: When conditions allow, the Vehicle will inform the Driver via the visual cockpit display that HAS is available to use. On seeing that HAS is available, the Driver activates the feature by pressing the HAS button. The system will take a short time to initialise. HAS will then invite the Driver to activate ACC. This is achieved via a momentary

ACTOR	ROLE	INTEREST / CONCERN
Driver	Uses driver inputs to control the HAS system	Activates HAS when required via the SET + or RESUME buttons. Is able to cancel the feature by pressing either the CANCEL button, or the brake pedal. Incremental changes to vehicle speed can be made by pressing SET + or SET -. Headway to the vehicle ahead may be increased or reduced. May momentarily override *Set Speed* by pressing the accelerator pedal while the feature is active.
Vehicle Environment	Dictates when it is safe for the driver to use HAS	Conditions may exist where it would be inappropriate for the driver to use the feature. For example, on a windy narrow road with no lane markings or in conditions like fog when vehicle sensors may fail to detect nearby objects.
Vehicle Manufacturer	Calibrates the HAS response	Requires vehicle behaviour, refinement and economy to be the same during HAS operation as normal driving. Requires the feature to behave consistently, thus promoting customer acceptance, safety and trust.

Table 7: HAS actor and stakeholder interests

press of the SET - button. The Vehicle will now maintain its speed or headway to the vehicle in front, and its position in lane. This may be the centre of the lane, but for travelling through a bend, or because the vehicle ahead takes up a non-centre line position, the *Reference Trajectory* may be offset from the lane centre.

The driver may use SET + and SET - to adjust the current vehicle *Set Speed*, or may press the accelerator pedal to momentarily override the current *Set Speed*. This will cancel HAS. The feature can be reenabled by pressing the SET - button. Additionally, the Driver may return to manual control at any time by pressing the CANCEL button, by pressing the brake pedal, or by using the vehicle's turning indicators.

In certain circumstances the vehicle may require the Driver to resume manual control, perhaps when the preceding vehicle makes an emergency stop, or when an adjacent vehicle cuts into the space ahead. In such cases an audible and visual warning will be given, and the feature will cancel. If detection of the Driver's hands on the steering wheel is lost, the Vehicle will warn the Driver, and if the situation persists will cancel HAS, returning the vehicle to manual driving.

PRE-CONDITIONS: The Vehicle is being driven on the highway, with the vehicle speed and lateral position being controlled by the Driver manually.

TRIGGER: Using GPS position and map data the system determines that the Vehicle has entered a designated highway.

B.2.2 *Requirements*

REQ 1: The CANCEL button or brake pedal inputs shall always take precedence over other HAS driver inputs, thus ensuring that the Driver is always able to cancel the automation if required.

REQ 2: If the system senses that the Driver is applying an overriding torque to the steering hand wheel then the system shall cancel.

REQ 3: If the system detects that the Driver has selected the direction indicators then HAS shall cancel. Note: ACC remains active.

REQ 4: An audible and visual *Driver Steering* warning shall be used to warn the Driver when detection of the Driver's hands on the steering wheel has been lost.

REQ 5: An audible and visual *Driver Override* warning shall be used to indicate to the Driver when a transition back to manual control is required.

B.2.3 *Use case: main flow*

1. The Vehicle indicates to the Driver via the visual cockpit display that HAS is available for use.
2. The Driver presses SET- button [*Alternative Flow 1: Press SET-*] to active ACC.
3. The vehicle's current speed is saved as the current *Set Speed* and the Vehicle begins longitudinal control.
4. The Driver then presses the HAS button to indicate that they wish to activate Highway Assist.
5. The Vehicle confirms that the lane boundaries are being detected before beginning to control the Vehicle's position in lane.
6. The Vehicle illuminates the visual cockpit display to inform the Driver that HAS is now controlling the vehicle's lateral position in lane.
7. Providing the Driver remains attentive [*Alternative Flow 2: Hands on Wheel Not Determined*] the Vehicle will maintain the correct vehicle position in lane, speed and proximity to other vehicles using the following [*Alternative Flow 3: Driver Take Back*].
    a) Information received from line-of-sight sensors
    b) GPS position information and map data

8. The Driver monitors the Traffic Environment and may:
   a) Change the *Set Speed* using SET+ [*Alternative Flow 4: Press SET+*] or SET- [*Alternative Flow 1: Press SET-*] buttons,
   b) Momentarily override HAS by moving the steering wheel, or by using the vehicle's direction indicators [*Alternative Flow 5: Lateral Override*],
   c) Cancel HAS and ACC by pressing the CANCEL button, or by pressing the brake pedal [*Alternative Flow 6: Cancel Automation*] as required.
9. The Vehicle remembers the last *Set Speed* for the remainder of the drive cycle.

POST-CONDITIONS HAS is inactive, the *Set Speed* is stored in memory, and the Driver controls the vehicle manually.

B.2.4 *Alternative flow 1: press SET -*

DESCRIPTION Pressing SET - when the system is inactive reactivates the system and sets the *Set Speed* to the current vehicle speed. While the system is active, pressing SET - will decrement the current *SET Speed*.
At Step 2:

1. If a *Set Speed* is stored the Vehicle activates ACC and rejoins the main flow at Step 3.
2. Else the current vehicle speed is stored as the new *Set Speed*, the Vehicle activates ACC and rejoins the main flow at Step 3.

At Step 8:

1. With ACC already active, the Vehicle decrements the *Set Speed* by a predetermined amount, and rejoins the main flow at Step 8.

B.2.5 *Alternative flow 2: hands on wheel not determined*

DESCRIPTION With HAS being classified as an SAE Level 2 system, the Driver remains responsible for vehicle safety while the automation is active. Therefore, if the Vehicle detects that the Driver has removed their hands from the steering wheel audible and visual *Driver Steering* warnings will be given. If the Driver fails to return their hands to the steering wheel the magnitude of the visual and audible warnings will escalate. If the situation continues to persist then the automation is cancelled.
At Step 7:

1. The Vehicle issues an audible and visual *Driver Steering* cockpit warning reminding the Driver to return their hands to the steering wheel.
2. If the Vehicle detects that the Driver's hands have returned to the steering wheel then the use case rejoins the main flow at Step 7.
3. A second audible and visual *Driver Steering* cockpit warning is issued, informing the Driver that if they do not return their hands to the steering wheel then HAS will disable.
4. If the Vehicle detects that the Driver's hands have returned to the steering wheel then the use case rejoins the main flow at Step 7.
5. Else a final audible and visual *Driver Steering* warning issued and the automation cancels.
6. The use case ends.

POST-CONDITION HAS is inactive, the previous *Set Speed* is retained in memory for the remainder of the drive cycle, and the Driver controls the vehicle manually.

B.2.6 *Alternative flow 3: driver take back*

DESCRIPTION Under certain circumstances HAS may request the Driver to take back full vehicle control. This could be due to interactions with other road users (e.g. the vehicle ahead makes an emergency stop, a vehicle from an adjacent lane cuts in sharply) or it could be that the Vehicle determines that it is no longer safe to operate HAS. This could be because GPS and map data indicates that the vehicle is no longer being driven on a dual carriageway (i.e. no longer in the ODD), or the GPS and map data is not being received, or a forward facing sensor has become blocked or has failed.
At Step 7:

1. The Vehicle provides the audible and visual *Driver Override* warning to inform the Driver that HAS is no longer available.
2. The feature disables and the use case ends.

POST-CONDITION HAS is inactive, the previous *Set Speed* is retained in memory for the remainder of the drive cycle, and the Driver controls the vehicle manually.

B.2.7 *Alternative flow 4: press SET +*

DESCRIPTION Pressing SET + while HAS is active increments the *Set Speed* by a calibrated amount.
At Step 8:

1. If HAS is active then the Vehicle increments the *Set Speed* by a predetermined amount, and rejoins the main flow at Step 8.

B.2.8 *Alternative flow 5: lateral override*

DESCRIPTION While HAS is active the Driver may wish to momentarily override lateral control – perhaps to change lanes or to overtake a slower vehicle. Using the vehicle's direction indicators or moving the steering wheel will cause lateral control to be suspended. ACC will remain active. Once the vehicle is again being driven in lane lateral control will reenable.
At Step 8:

1. The Driver uses the vehicle's direction indicators or simply moves the steering wheel to indicate that they wish to suspend HAS and steer manually.
2. The Vehicle deactivates lane centring and indicates to the Driver via the visual cockpit display that HAS is no longer active.
3. The Driver completes their steering manoeuvre and returns the vehicle to the lane centre.
4. Identifying that the steering manoeuvre is complete, the Vehicle reactivates lateral control and informs the Driver via the visual cockpit display that HAS is active.
5. The use case rejoins the main flow at Step 7

B.2.9 *Alternative flow 6: cancel automation*

DESCRIPTION The Driver may regain manual vehicle control at any point by pressing either the CANCEL button or the brake pedal. In either case the *Set Speed* remains stored in memory for the remainder of the drive cycle.
At Step 8:

1. The Driver presses either the CANCEL button or the brake pedal indicating that they wish to cancel HAS.

2. The Vehicle deactivates the automation and remembers the *Set Speed* for the remainder of the drive cycle.
3. The use case ends.

POST-CONDITIONS HAS is inactive, the *Set Speed* is stored for the remainder of the drive cycle, and the Driver controls the vehicle manually.

B.3 AUTOPILOT

B.3.1 *Overview*

The Autopilot feature is designated as an SAE Level 4 system [141]. The system is designed to allow the Driver to completely relinquish control and responsibility for the DDT and for vehicle safety while the system is operational within the ODD. Unlike lower levels of automation, the Driver is not required to remain attentive and potentially be required to take back control to maintain vehicle safety.

There are no SAE Level 4 automated systems currently in production. The below use case has been written to reflect the behaviour of Tesla's Navigate on Autopilot, albeit written as an SAE Level 4 system rather than an SAE Level 2 / 3 production system [158].

GOAL: To take full control of the DDT on the Motorway network, in areas where feature activation is permitted. While active the feature does not require the Driver to remain vigilant, allowing them to engage in secondary non-driving related tasks (e.g. reading).

BRIEF DESCRIPTION: Autopilot is designed to work on predefined sections of the road network. These are multi-lane carriageways having a central crash barrier between the lanes and on-coming traffic. Before starting their journey the Driver enters the required destination into the navigation system. If the route selection includes an Autopilot

ACTOR	ROLE	INTEREST / CONCERN
Driver	Indicates when they wish to use Autopilot and sets desired cruising speed	When the system indicates that Autopilot is available the Driver lifts the cruise control stalk twice to relinquish control, allowing the vehicle to control the DDT. The Driver may adjust the vehicle's target cruising speed up or down using the rotary dial on the steering wheel.
Vehicle Environment	Dictates when it is safe for the driver to use Autopilot	The feature will only activate once the vehicle is on a designated part of the road network. With the Driver being permitted to engage in non-driving tasks while Autopilot is active, the system retains responsibility for vehicle safety under all conditions.
Vehicle Manufacturer	Develops the Autopilot feature	Requires Autopilot to exhibit a defensive driving style and to behave consistently, thus promoting customer acceptance, safety and trust.

Table 8: Autopilot actor and stakeholder interests

permissible route, then the Driver can indicate their wish to use Autopilot by selecting it on the navigation screen. Assuming that the vehicle is not already in the ODD, the Driver will need to begin driving manually. Once Autopilot detects that it has reached a designated section of the road network, the system will indicate to the Driver that Autopilot is now available. The Driver simply lifts the cruise control stalk twice, which enables Autopilot.

Once active, Autopilot is capable of merging into the traffic flow (from the Motorway on-ramp), change lane when needed, and then before the required Motorway exit is reached, manoeuvre into the correct lane and exit the main carrigeway. Before the vehicle reaches the end of the ODD the system will indicate to the Driver (via visual and audible warnings) that they need to resume control. The Driver lifts the cruise control stalk once and resumes manual control.

PRE-CONDITIONS: Before departure the Driver has selected to navigate a route that includes section(s) of road where Autopilot use is permitted and has selected the option to use the feature when it becomes available. The Vehicle is receiving all information necessary (e.g. GPS) to enable Autopilot, all on-board diagnostics have successfully completed and no faults have been identified.

TRIGGER: The vehicle is being driven manually towards the highway, and using GPS position and map data the system determines that the Vehicle is entering a Motorway on-ramp where Autopilot use is permitted.

B.3.2 *Requirements*

REQ 1: The Driver shall acknowledge their wish to relinquish DDT control by lifting the cruise control stalk twice.

REQ 2: An audible and visual *Driver Take-Back Control* warning shall be used to indicate to the Driver when a transition back to manual control is required.

REQ 3: The Driver will indicate to the system that they are ready to resume manual control of the DDT by lifting the cruise control stalk once.

REQ 4: If the Driver fails to regain manual DDT control before the end of the ODD is reached then the system must manoeuvre the vehicle to a position of safety, stop and secure the vehicle.

B.3.3 *Use case: main flow*

DESCRIPTION: The below describes the normal flow through the use case.

1. Having entered the Motorway on-ramp where Autopilot use is allowed, the system displays the Autopilot symbol informing the Driver that the feature is now available to use.
2. Seeing the Autopilot symbol appear on the display, the Driver lifts the cruise control stalk twice to indicate that they wish to activate Autopilot.
3. Autopilot takes over DDT control and indicates to the Driver that the feature is now active by colouring the Autopilot symbol blue.
4. The Driver removes their hands and feet from the driver controls, pushes back the seat and closes their eyes.
5. The Vehicle progresses down the on-ramp, with Autopilot controlling the Vehicle's speed and position in lane as appropriate for the speed and position of other traffic, the road geometry and the speed limit.
6. As the Vehicle approaches the end of the on-ramp Autopilot monitors the speed and position of other road users on the Motorway, adjusting the Vehicle's speed and position accordingly to safely merge the Vehicle onto the main carriageway.
7. The Vehicle proceeds along the Motorway, with Autopilot controlling the Vehicle's speed and position in lane as appropriate given the desired target speed selected by the Driver, traffic flow and density, the road

geometry and the speed limit. Autopilot continuously determines the most appropriate lane to achieve best progress towards the desired exit, changing lanes as required [*Alternative Flow 1: Change Lane*], while also continuously monitoring the following:

a) Presence of a collision risk [*Alternate Flow 2: Collision avoidance*]

b) GPS position information and high-definition map data loss [*Alternate Flow 3: Missing off-board data*]

c) Sensor redundancy loss (e.g. sensor element obscured) [*Exception Flow 1: Non-critical error*]

d) Critical failure affecting on-board line-of-sight sensor or system actuator integrity [*Exception Flow 2: Critical system failure*]

8. When the GPS position information indicates that the Vehicle is nearing (less than 1000 metres) the required Motorway exit, Autopilot moves the Vehicle into Lane 1, changing lanes as required [*Alternative Flow 1: Change Lane*].

9. Autopilot detects the beginning of the exit ramp and steers the Vehicle to leave the main carriageway.

10. Autopilot controls the Vehicle down the Motorway off-ramp at a speed appropriate to the road geometry, the speed limit and traffic flow, and issues an audible and visual *Driver Take-Back Control* warning, indicating to the Driver that they need to regain DDT control.

11. On hearing the audible chime and visual warning, the Driver returns their seat to the correct driving position, places their feet on the pedals and hands on the wheel.

12. Autopilot continues controlling Vehicle speed and position on the off-ramp and the use case ends when either the Driver regains control (by lifting the cruise control stalk once) and DDT control returns to the Driver, or the end of the ODD is reached [*Exception Flow 3: ODD End*].

POST-CONDITIONS Autopilot is inactive and the Driver controls the vehicle manually.

B.3.4 *Alternative flow 1: lane change manoeuvre*

DESCRIPTION While Autopilot is operational there will be times when the Vehicle needs to change lane. For example, approaching a slower vehicle, termination of the current lane ahead, traffic in the adjacent lane is moving more quickly. At any point during the journey Autopilot may request a lane change, either to the left or right. The main Autopilot use case makes the strategic lane change decisions and will not make an illegal lane change request (e.g. undertaking manoeuvre when it's not permitted). Consequently, the change lane functionality is responsible purely for the safe execution of the lane change manoeuvre. If the main use case wishes to move the Vehicle across more than one lane, it will issue multiple requests – with control returning to the main use case after each lane change.

PRE-CONDITION Autopilot makes a request to change lanes, either a lane change to the left or to the right: At Step 7 (lane change), Step 8 (approaching exit) or *Exception Flow 3: ODD End* Step 1 (moving towards hard shoulder).

1. Sensors monitor the adjacent lane into which the Vehicle will be moving, both to the side and behind, to determine the location and closing speed of nearby vehicles.
2. If a gap of sufficient length is identified in the adjacent lane, Autopilot activates the Vehicle's direction indicators and begins to modify the Vehicle speed to match that of the traffic in the adjacent lane. Else, if no gap is identified within a predetermined time period (longer for *Exception Flow 3*), the Vehicle remains in the current lane and the use case rejoins the main flow at Step 7, Step 8 or *Exception Flow 3: ODD End* Step 1.

3. Autopilot adjusts the Vehicle's lateral position to move it into the adjacent lane and cancels the Vehicle's direction indicators.
4. The use case rejoins the main flow at either Step 7, Step 8 or *Exception Flow 3: ODD Loss* Step 1.

B.3.5 *Alternative flow 2: collision risk*

DESCRIPTION While Autopilot is operational situations may arise where an object that poses a collision risk is identified in front of the Vehicle. To mitigate the collision risk Autopilot can either change lane, or slow down and stop before the object in path is reached. To maintain progress towards the destination and to minimise the risks of stopping in a 'live' lane, Autopilot will always opt to change lane where possible. The decision point (i.e. time distance to the detected object) when Autopilot takes the decision to brake rather than change lane is predefined – requiring potentially heavy but not emergency braking force.

PRE-CONDITION An object that represents a collision risk has been identified in the Vehicle's lane. At Step 7:

1. Using forward and reverse sensors Autopilot determines the distance to the object and the location, distance and speed of other road users – both those in the current lane and those in the lane into which the Vehicle will move.
2. While the time distance to the object remains above the predefined threshold:
    a) Autopilot attempts to identify a gap of sufficient length in the adjacent lane, into which the Vehicle can move.
    b) If a gap of sufficient length is identified in the adjacent lane, Autopilot activates the Vehicle's direction indicators, modifies the

Vehicle's speed to match the adjacent traffic's speed and moves into the adjacent lane.

c) Control returns to the main flow at Step 7.

3. Autopilot applies the brakes to slow and stop the vehicle with the hazard lights illuminated before the obstacle is reached, while issuing a *Driver Take-Back Control* warning informing the Driver that they need to regain DDT control.

4. The use case ends.

POST-CONDITIONS Autopilot is inactive, the Vehicle is stationary and DDT control is returned to the Driver.

B.3.6 *Alternative flow 3: missing off-board data*

DESCRIPTION While Autopilot is operational situations may arise where off-board GPS position and high-definition map data is on longer being received. Typically this occurs when the Vehicle travels through an area where radio reception is not possible. For example, a road tunnel. Autopilot holds internal map data which it correlates with objects identified by the on-board sensors, so is able to tolerate such data losses for a period of time. However, with the off-board data missing Vehicle position accuracy will reduce over time and so the Vehicle behaviour must be modified accordingly. Eventually reaching the point where the Driver will be required to regain DDT control to ensure Vehicle safety is maintained; at which point the Vehicle is deemed to have left the ODD.

PRE-CONDITION Off-board GPS position and high-definition map data is no longer being received. At Step 7:

1. An Internal Timer is started.

2. While the off-board data remains missing the Internal Timer increments and Autopilot calculates the Vehicle's 'dead-reckoning' position from Vehicle speed and current trajectory:

    a) If the Internal Timer < *Threshold 1* Autopilot maintains the Vehicle's speed and position in lane and will update Vehicle position if internal map roadside objects are identified by the on-board sensors.

    b) If *Threshold 1* < Internal Time > *Threshold 2* Autopilot issues the *Driver Take-Back Control* warning to inform the Driver that they need to regain DDT control while beginning to slow the Vehicle. Autopilot will update the current 'dead-reckoning' position if internal map roadside objects are identified by the on-board sensors.

    c) If the Driver lifts the cruise control stalk, to indicate that they are regaining DDT control, Autopilot relinquishes DDT control and the use case ends.

    d) If the Internal Time > *Threshold 2* then the end of the ODD is deemed to have been reached [*Exception Flow 3: ODD End*] and the use case ends.

3. Autopilot begins receiving off-board GPS and high definition map data and the use case rejoins the main flow at Step 7.

B.3.7 *Exception flow 1: non-critical error*

DESCRIPTION The Autopilot system incorporates redundant sensor technologies to maximise the system's perception capabilities. Consequently, if a sensor is obscured or fails Autopilot can continue to operate, albeit with performance limitations in place – e.g. the safe distance to other road users is increased, and the maximum Vehicle speed is reduced.

PRE-CONDITION An on-board sensor has either become obscured or has failed. At Step 7:

1. Having detected that an Autopilot sensor has been obscured or has failed, the parameters within Autopilot's safe operating envelope are modified (e.g. maximum speed reduced, headway distance increased, lane change manoeuvres no longer permitted) according to the nature of the sensor loss.
2. For an obscured sensor, the performance modifications persist while the obstruction persists.
3. For a failed sensor, the performance modifications persist until the system is reset by a Service Technician.
4. Control returns to the main flow at Step 7.

B.3.8 *Exception flow 2: critical system failure*

DESCRIPTION A redundant actuator strategy is also deployed, allowing Autopilot to continue operating with an actuator failure present. However, with the loss of a second actuator having potentially catastrophic effects on Vehicle safety, Autopilot operating with a single actuation failure is not allowed to persist. Consequently, on detection of an actuator fault Autopilot will notify the Driver that they need to resume DDT control. If the Driver fails to regain DDT control within a predefined time the Vehicle is deemed to have left the ODD and will be brought to a stop in a safe place.

PRE-CONDITION A Vehicle actuator used by Autopilot has failed. At Step 7:

1. Having detected that an Autopilot actuator has failed, the parameters within Autopilot's safe operating envelope are modified (e.g. maximum speed reduced, headway distance increased, lane change manoeuvres no longer permitted) according to the nature of the failure detected.
2. Autopilot issues an audible and visual *Driver Take-Back Control* warning, indicating to the Driver that they need to regain DDT control.

3. On hearing the audible chime and visual warning, the Driver returns their seat to the correct driving position, places their feet on the pedals and hands on the wheel.
4. Autopilot continues controlling Vehicle speed and position and the use case ends when either the Driver regains control (by lifting the cruise control stalk once) and DDT control returns to the Driver, or after the predefined time period has elapsed without the Driver regaining DDT control [*Exception Flow 3: ODD end*].

POST-CONDITIONS Autopilot is inactive and the Driver controls the vehicle manually.

B.3.9 *Exception flow 3: ODD end*

DESCRIPTION In situations were the Vehicle is deemed to have left the ODD while Autopilot has DDT control, the Vehicle must be brought to a stop in a safe location with hazard lights flashing, the parking brake applied, an emergency E-Call sent and the Vehicle's doors unlocked.

PRE-CONDITION The Vehicle is deemed to have left the ODD while Autopilot is undertaking DDT control. At Step 12 (end of designated Motorway) or *Exception Flow 2: Critical System Failure* Step 4 (duration operating without redundant actuators exceeded)

1. If the Vehicle is not currently in the lane adjacent to the hard shoulder, then Autopilot will control the Vehicle into that lane [*Alternative Flow 1: Lane Change*].
2. Once in the lane adjacent to the hard shoulder, the Vehicle's hazard lights are illuminated and Autopilot begins to slow the Vehicle.

3. Using a combination of on-board sensors and map data, off-board high-definition map and GPS position data, Autopilot identifies a safe location to stop the Vehicle.
4. Autopilot manoeuvres the Vehicle into the safe position identified, applies the parking brake, sends an emergency E-Call, unlocks the Vehicle's doors, and the use case ends.

POST-CONDITIONS Autopilot is inactive, the Vehicle is stationary and switched off.

C

HIGHWAY ASSIST (LANE CENTRING FUNCTION) CASE STUDY

C.1 INTRODUCTION

Through 2018 and 2019 NHTSA published a number of reports as part of the "Automotive Electronics Reliability Research Program". This programme had a number of goals focused on developing the methodologies, processes and best practice for an Automated Lane Centring (ALC) vehicle feature and the the foundation systems (such as power steering) needed to support the lane centring function. This has resulted in a number of reports being published, covering the central system and these foundation systems. Each report has included a hazard analysis section deploying three analysis techniques: HAZOP, Functional Failure Mode Effects Analysis (FMEA) and STPA. Comparing the STPA analysis from the 'core' ALC report [21] allows the outcomes from applying the EVCM and *Shared Control* STPA method proposed by this research to be compared directly with the outcomes from NHTSA independent research team.

This chapter follows the four step shared control hazard analysis method (see Figure 12, page 110) first described in Chapter 7. Section C.2 describes the ALC system as described in the NHTSA report [21], together with the system diagrams presented in that report. The behavioural competencies applicable to ALC are then identified. Having identified the applicable behavioural competencies a behavioural competency interaction diagram is drawn. From this the control actions that are shared between the automation and human driver are identified. Finally, the CSD is drawn for ALC – created by populating the EVCM with the pertinent behavioural competencies identified. Section C.3.1 then uses the STPA guide words to identify the UCAs for ALC before analysing

those UCAs in the context of shared control to identify applicable loss scenarios. Like the NHTSA report, appropriate safety constraints and safety requirements are suggested as part of the analysis. For a full implementation, the loss scenarios identified in Section C.3.1 would help inform the test strategy. The NHTSA report includes a "Performance parameters and test scenarios" section, which is compared with the outcomes from this analysis in Section C.4. The chapter concludes with a summary.

C.2 BEHAVIOURAL COMPETENCY SELECTION

C.2.1 *Automated lane centring*

Functional Safety Assessment of an Automated Lane Centering System [21] describes a generic ALC function which could be used across all SAE automation levels [141]. The job of the ALC function is to provide continuous lateral vehicle control that keeps the vehicle on a reference trajectory. In the context of an SAE Level 2 vehicle feature such as Highway Assist System (HAS), the reference trajectory should be thought of as the 'ideal' safe lane position for the vehicle. For SAE Levels 3 and 4 the reference trajectory would be generated by the path planning function. In either case, it should be noted that the safe reference trajectory may not be the lane centre. For example, when there is a road curvature or when passing a wide vehicle in the adjacent lane it may be more appropriate to operate the vehicle off-centre.

To help bound the problem the SAE Level 2 automation description from the NHTSA report is used here. Typically, commercially available AD features, such as Alfa Romeo's HAS, combine this SAE Level 2 type lane centring functionality with ACC longitudinal control functionality to provide the production vehicle feature [6, 25]. For reference, a use case description of the full HAS functionality (i.e. ACC with ALC) appears in Section B.2, page 231.

The NHTSA report includes the generic block diagram for ALC (see Figure 31) together with a description of the functions that each diagram 'block' fulfils.

This identifies the following behavioural competencies: *maintain position in lane, ODD detection* and *operator engagement*. Considering the requirement to maintain the safety of the vehicle within the ODD identifies the further strategic and manoeuvring level behavioural competencies of: *perform surveillance, comply with the rules, avoid obstacles* and *right-of-way decision*. Arguably operating such a system in a motorway context will also require the *navigate temporary conditions* behavioural competency. However, with the NHTSA report not discussing the system behaviour under temporary restrictions (e.g. road works, accident resulting in lane closure) this behavioural competency is excluded from this analysis.

The identified behavioural competencies are described below, with the driver / automation responsibility split being described in Table 9 (page 257):

- Strategic Level Competencies
  - PERFORM SURVEILLANCE: A behavioural competency undertaken by the human driver, requiring them to understand the factors in the vehicle's environment that could influence the vehicle's safe progress. This includes making changes to the vehicle operation (including the use of automation) necessary to maintain safe progress.
  - COMPLY WITH THE RULES: In the context of ALC operation on a multi-lane carriageway, this will involve complying with the rules of the road (e.g. minimum permissible speed) and using information and advice from road-side signage (e.g. speed limit) to select the appropriate speed for the vehicle. A vehicle feature like HAS may be able to adjust the vehicle target speed to reflect the speed limit (i.e. when speed limit changes are embedded in map data), but oversight by the human driver may be needed in other context.
  - RIGHT-OF-WAY DECISION: Involves sensing traffic control devices, signage and infrastructure detail to determine the correct vehicle path. This might include identifying when the current vehicle lane ends or when traffic is required to merge. For an SAE Level 2

system this behavioural competency will be largely undertaken by the driver.

- Manoeuvring Level Competencies
  - AVOID OBSTACLES: Involves identifying objects in the vehicle's path that present a collision risk. When ALC is used in conjunction with ACC, this behavioural competency will be partially covered by the ACC detecting preceding vehicles and modifying the vehicle's speed accordingly. In parallel the driver is also expected to continuously monitor for collision risks (e.g. adjacent vehicle cutting in, object falling off preceding vehicle, pedestrians) and manage the situation accordingly.
  - FOLLOW OTHER VEHICLES: The aspect of longitudinal vehicle control automated in part by ACC. Although the ACC system will be able to maintain the vehicle's safe *headway* to the preceding vehicle, the driver may still need to intervene. For example, if an adjacent vehicle cuts-in or the preceding vehicle performs an emergency stop.
  - MAINTAIN VEHICLE SPEED: The other behavioural competency need for vehicle longitudinal control. Used when no preceding vehicle is detected. Again, in the context of HAS this would be performed by the ACC function.
  - MAINTAIN POSITION IN LANE: While active ALC must maintain the vehicle on the *reference trajectory*. This involves determining the lane boundaries, but also identifying the position of other vehicles on the road. With the relative position to other vehicles potentially causing ALC to modify the *reference trajectory* to maintain a safe separation from other road users. For example, if the subject vehicle passes a truck carrying a wide load.
- Control Level Competencies
  - PERFORM LONGITUDINAL (BRAKING) CONTROL: Longitudinal vehicle control task that decelerates the vehicle to achieve the target (negative) longitudinal acceleration defined by behaviours such as *follow*

other vehicles. Typically, ACC does not have full authority over braking, so in scenarios requiring a full force brake application (e.g. an emergency stop) the driver will be required to intervene to achieve the required braking force.

  - PERFORM LONGITUDINAL (ACCELERATION) CONTROL: Longitudinal vehicle control task that accelerates the vehicle to achieve the target longitudinal acceleration determined by the manoeuvring level behavioural competencies (e.g. *follow other vehicles, maintain speed*).
  - PERFORM LATERAL CONTROL: Controls the vehicle's steering system to achieve the lateral acceleration needed to achieve the reference trajectory set by *maintain position in lane*.

Not directly related to DDT control, the NHTSA report highlights two further applicable competencies: *Assess driver engagement* and *transition of control*:

- Automation use
  - ASSESS DRIVER ENGAGEMENT: Being an SAE Level 2 system (i.e. the driver is responsible for vehicle safety) there is a requirement to continually monitor that the driver remains vigilant. Typically this involves checking that the driver's hands remain on the wheel and monitoring the driver's gaze (to access their level of attentiveness).
  - TRANSITION OF CONTROL: Covers the driver operated controls and audible and visual warnings used to communicate system status information to the driver. These controls allow the driver to activate, deactivate and override the system. While the audible and visual warnings allow the system to warn the driver when the automation needs to deactivate – for example, because a system fault has occurred.

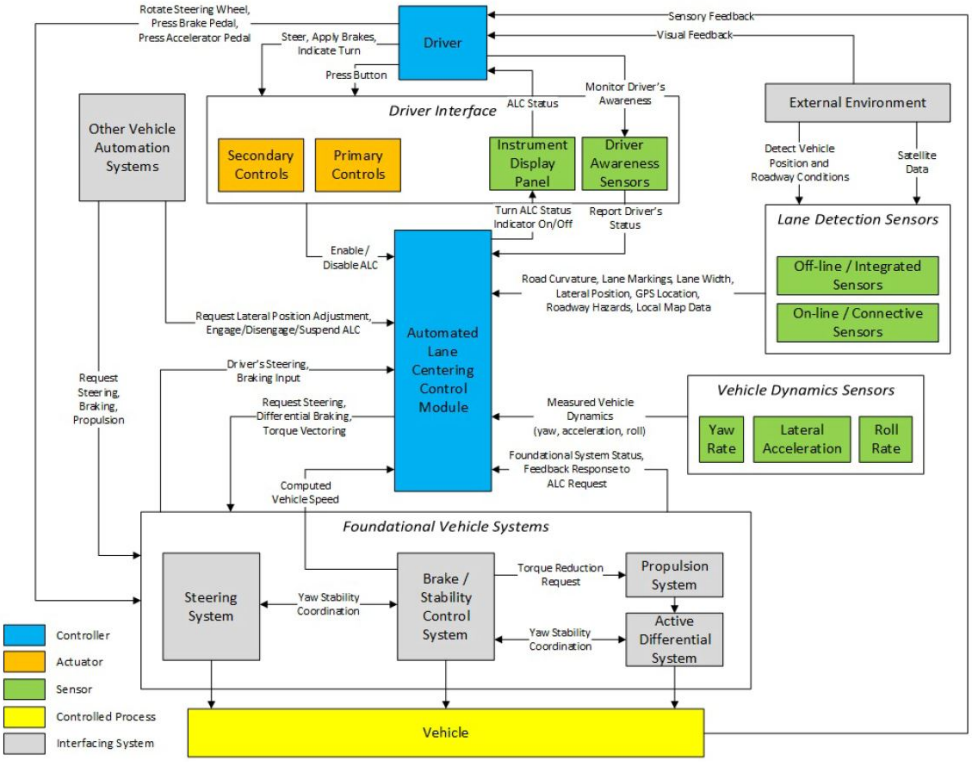


Figure 31: The block diagram of a generic ALC system (from [21])

C.2.2 *Behavioural competencies for safety*

The HAS behavioural competencies discussed above are considered in the context of the behavioural competency taxonomy (see Appendix A, page 210). Some behavioural competencies are considered out-of-scope of the anticipated ODD, others are considered necessary to maintain safety. This analysis is captured in Table 9.

Some behavioural competencies are considered out-of-scope for HAS. For example, *start the vehicle, make the vehicle safe* and *deal with different road types*. Others, such as the control level competencies are solely the responsibility of the automation – i.e. *hold the vehicle stationary, pull away from standstill, perform lateral (steering) control, perform longitudinal (accel.) control, perform longitudinal (decel.) control. Perform surveillance* is the one behavioural competency considered the sole responsibility of the human driver. The responsibility for the remaining

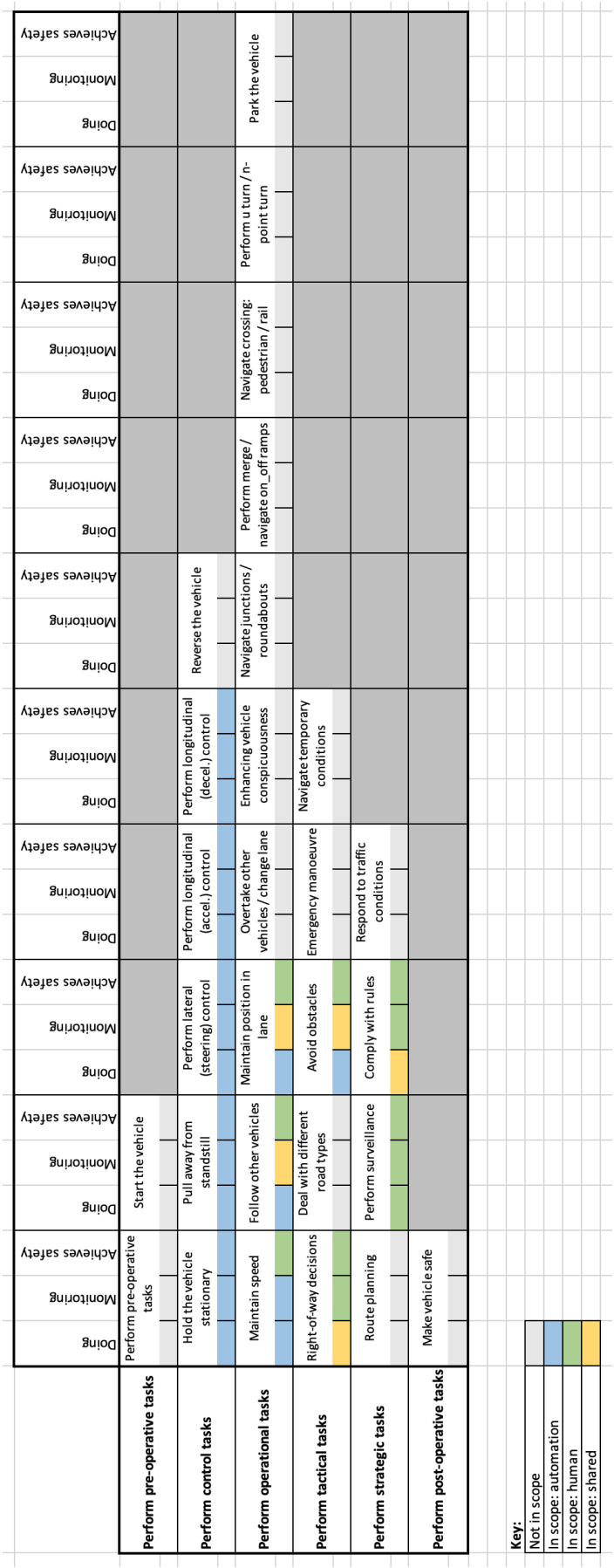


Table 9: HAS actor responsibility behavioural competency comparison: who is doing, who is monitoring and who is responsible for vehicle safety?

behavioural competencies are considered and are discussed in more detail below.

Maintain speed is largely the responsibility of the automation. However, achieving safety for the competency remains with the driver while the automation is active. For example, external factors may dictate a reduction in vehicle speed, which would require the driver to modify the *Set Speed* accordingly. Consequently, 'doing' and 'monitoring' are designated as the automation's responsibility (coloured blue), while 'achieve safety' is designated as the human's responsibility (hence coloured green).

Following other vehicles, *maintain position in lane* and *avoid obstacles* see the responsibility split between the human and the automation. In each case the 'doing' is considered the responsibility of the automation (coloured blue). 'Monitoring' is considered to be a shared responsibility (coloured yellow). That is, the automation has some capability to monitor the behavioural competency, however it is anticipated that there will be situations where the human driver will need to monitor the behaviour. For example, heavy braking scenarios where both *follow other vehicles* and *avoid obstacles* would require the driver to intervene to achieve the required deceleration rate to avoid a collision.

Two behavioural competencies have the 'doing' designated as shared (coloured yellow) and 'monitoring' and 'achieve safety' designated as being the human's responsibility. They are *right-of-way decision* and *comply with rules*. For example, the automation may receive speed limit information, via map data, allowing it to reduce vehicle speed when required. However, if a temporary speed limit is in place or an error exists in the map data, then the driver would need to manually modify the current vehicle speed.

Figure 32 depicts the interaction between the behavioural competencies discussed. With all control level behavioural competencies being the responsibility of the automation (i.e. no interaction between the actors) these behavioural competencies are omitted from the behavioural competency interaction diagram to reduce clutter. The diagram shows the relationship between the behavioural competencies needed to maintain safety.

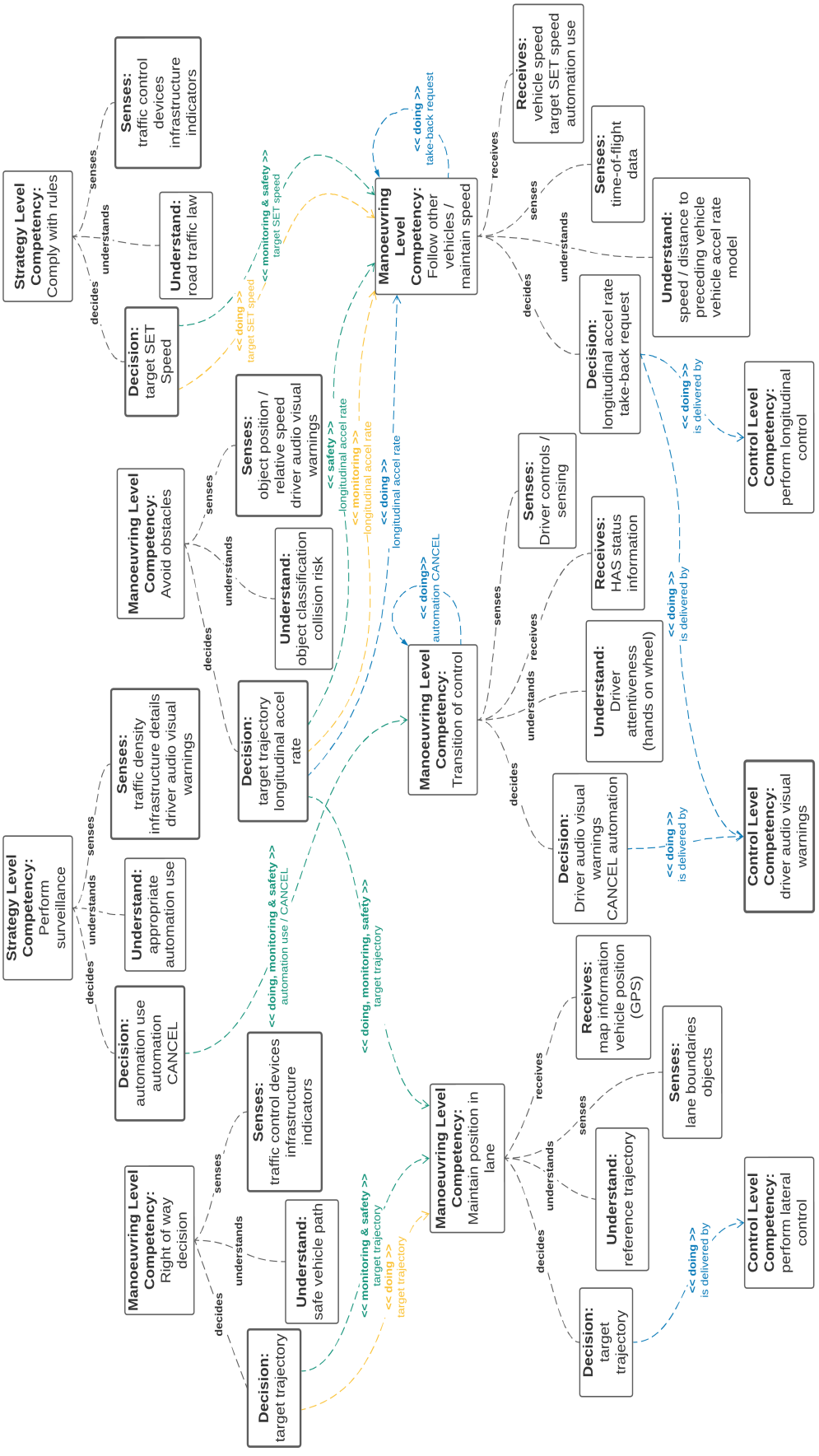


Figure 32: HAS behavioural competency interaction diagram

C.3 THE ANALYSIS OF AUTOMATED LANE CENTRING

The above behavioural competency discussion has considered all behavioural competencies applicable to the HAS vehicle feature. To aid the comparison with the NHTSA report [21] the remainder of the analysis focuses specifically on lateral control aspects of the vehicle feature.

C.3.1 *Define the purpose of the analysis*

The NHTSA Report considers the hazards associated with the loss of lateral vehicle control, and as such defines the following hazards: insufficient / excessive lateral adjustment resulting in lane departure, unexpected loss of ALC, improper transition of control between ALC and the driver, and ALC behaviour impedes other vehicle systems. The analysis of HAS identifies a similar hazard list (see Table 10). The hazards identified by this analysis are arguably more 'all encompassing' for two reasons: firstly, considering the DDT as a shared activity means an *improper transition of control* or *ALC interfering with other systems* are treated as hazard causes, rather than hazards in their own right. Secondly, with HAS also controlling longitudinal vehicle movement, hazards related to longitudinal control are included. Not considered by the NHTSA Report is the loss of vehicle stability, which given that ALC is controlling vehicle yaw moment and lateral acceleration does seem a critical omission.

C.3.2 *Control structure diagram and UCAs*

The CSD for HAS appears in Figure 33. The CSD has been created by populating the EVCM with the identified behavioural competencies, control actions and the information fed back to the identified behavioural competencies.

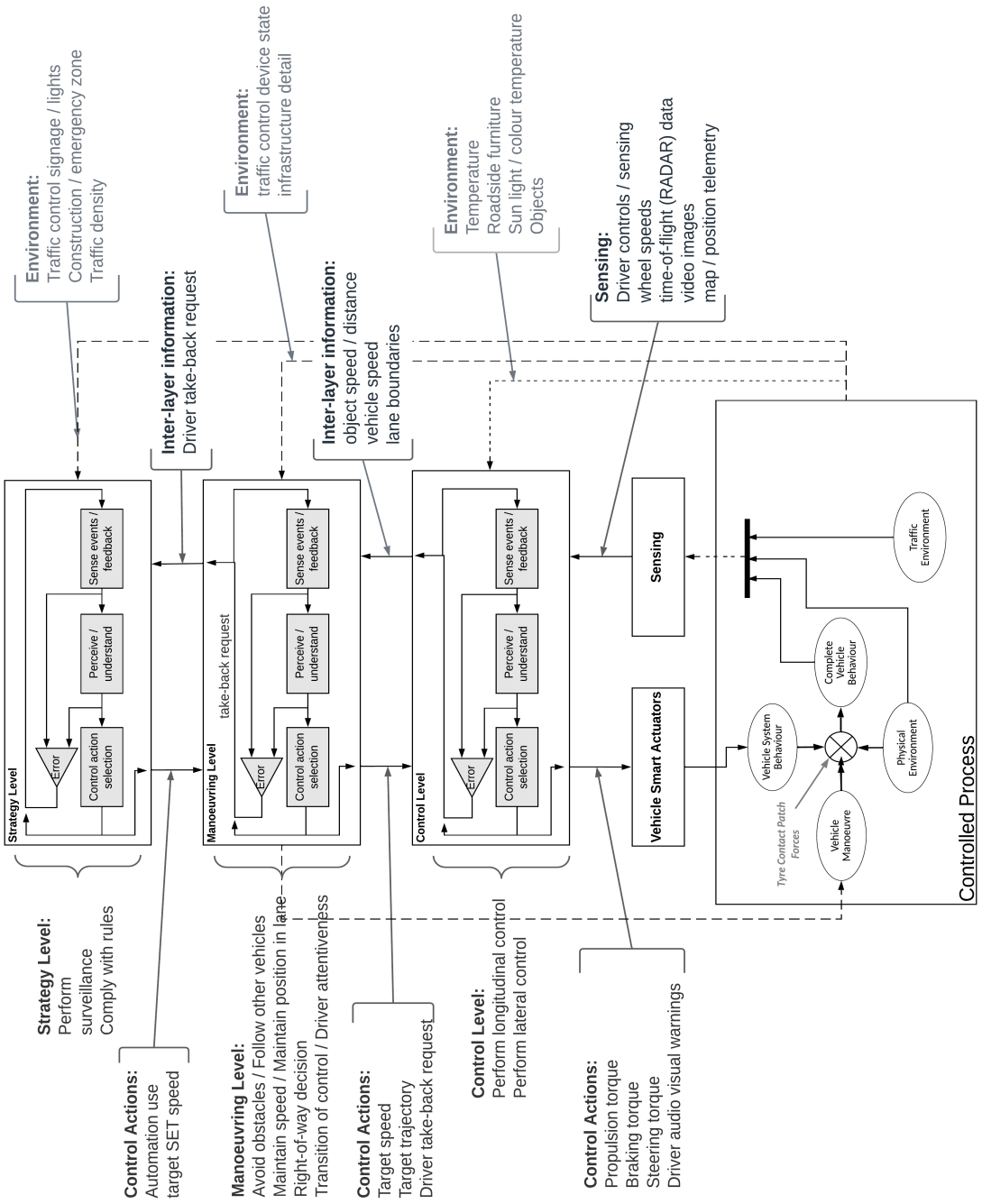


Figure 33: CSD for HAS

Stakeholders
The Driver of the vehicle
Vehicle occupants
Other road users
Insurance companies
Vehicle manufacturer

Losses	
L1	Loss of life or injury to people
L2	Loss of or damage to vehicle
L3	Loss of or damage to objects outside the vehicle
L4	Loss of consumer confidence or sales
L5	Financial loss

Hazards		Potential Losses
H1	Vehicle does not maintain minimum safe distance to preceding / adjacent vehicles	L1, L2, L3, L4, L5
H2	Vehicle fails to stop for / avoid object in path	L1, L2, L3, L4, L5
H3	Succeeding vehicles cannot maintain minimum safe distance due to vehicle's deceleration rate	L1, L2, L3, L4, L5
H4	Vehicle manoeuvre leads to vehicle instability / loss of stability	L1, L2, L3, L4, L5

Table 10: Identified stakeholders, losses and hazards for a Highway Assist System (HAS)

C.3.3 *Loss scenarios*

The below tables show the UCAs, functional insufficiencies, triggering conditions and safety constraints for each control action identified. As per Step 3 of the *'Classic'* STPA process, the UCAs have been identified by applying the guide words shown at the top of each column. Being an SAE Level 2 system [141], responsibility (i.e. the 'doing', 'monitoring' and 'achieves safety') for many behavioural competencies is shared between the driver and the automation. Consequently, each UCA is analysed in the context of each applicable behavioural competency. For example, analysing the control action *target trajectory* with the guide word *not providing causes hazard* is applicable to the behavioural competencies *maintain position in lane, avoid obstacles* and *right-of-way decision*.

Asking the loss scenario questions (see Figure 17, Chapter 9, page 137) then identifies the applicable loss scenarios from the UCAs. Finally, the safety requirements are written to mitigate both the UCAs and loss scenarios identified. Given that the focus here is lateral control, the loss scenarios relating to *target trajectory* UCAs are discussed in more detail below.

Note to Tables 13, 14, 15: The numbers and letters in brackets next to each loss scenario identify the loss scenario question (from Figure 17 page 137) originally identifying them.

C.3.3.1 *1 – Unsafe controller behaviours or inadequate feedback*

The analysis uncovers a number of behavioural competencies where *unsafe controller behaviour (1a)* (from the automation, the driver, or both) result in an incorrect *target trajectory* determination. In the case of the automation, UCAs occur either because an automation fault leads to an incorrect calculation, or because the automation does not cancel when the driver attempts to override the system via the steering wheel. The driver enabling the automation when driving conditions make it unadvisable to do so is another potential cause of an UCA. Finally, considering *unsafe controller behaviour* in the context of two behavioural competencies together (namely, *right-away-decision* and *maintain position in lane*) identifies a scenario where control inputs by the driver and automation together could result in an accumulative effect. For example, the automation making a positive adjustment to *target trajectory*, to maintain the vehicle's position in lane, while the driver also makes a positive steering input to change lane. The net effect of a large *target trajectory* change results in a vehicle instability (see UCA8 Table 15).

The analysis of UCA1 and UCA4 also highlights loss scenarios where the controller receives incorrect feedback or information. For example, incorrect vehicle location determination due to errors in the HD map data or the HD map data being out-of-date is a potential cause of both UCA1 and UCA4. As is the SOTIF triggering condition of poor lane makings.

C.3.4 *2 – Control actions improperly executed or not executed*

Questions related to *scenarios involving the control path (2a)* and *scenarios related to the controlled process (2b)* identifies further loss scenarios. The analysis of UCA2 identifies a loss scenario where the driver's input is insufficient to override the

Control Action	Unsafe Control Actions			
	Not providing causes hazard	*Providing causes hazard*	*Provided but timing wrong*	*Provided but duration wrong*
Target trajectory	UCA1: *maintain position in lane* does not provide target trajectory changes to maintain vehicle's central position in lane [H1]	UCA4: *maintain position in lane* provides a target trajectory change when no path deviation is required [H1]	Covered by *providing causes hazard*	Covered by *providing causes hazard* and *not providing causes hazard*
	UCA2: *avoid obstacles* does not provide target trajectory override when object in path [H2]	UCA5: *maintain position in lane* provides target trajectory (centring) while *avoid obstacles* is also changing target trajectory to avoid object in path [H2]		
	UCA3: *right-of-way decision* does not provide target trajectory override when lane change required [H2]	UCA6: *maintain position in lane* provides target trajectory (centring) while *right-of-way decision* is modifying target trajectory to change lane [H2]		
		UCA7: *avoid obstacles* provides target trajectory change while *maintain position in lane* is also changing target trajectory (i.e. accumulative effect) [H4]		
		UCA8: *right-of-way decision* provides target trajectory change while *maintain position in lane* is also changing target trajectory (i.e. accumulative effect) [H4]		

Table 11: UCAs for the HAS *Target Trajectory* control action

Control Action	Unsafe Control Actions			
	Not providing causes hazard	*Providing causes hazard*	*Provided but timing wrong*	*Provided but duration wrong*
Longitudinal acceleration / deceleration	not applicable to *maintain speed*	UCA11: *maintain speed* provides a large longitudinal deceleration with another vehicle close behind [H3]	Wrong timing already covered	Wrong duration not applicable to *maintain speed*
	UCA9: *follow other vehicles* does not provide longitudinal deceleration when preceding vehicle slows [H1]	UCA12: *maintain speed* provides a large longitudinal deceleration on a low mu surface [H4]	UCA17: *avoid obstacles* provides longitudinal deceleration too late when driver take-back control requested [H1, H2]	Wrong duration same as *not providing* for *follow other vehicles* and *avoid obstacles*
	UCA10: *avoid obstacles* does not provide longitudinal deceleration change when object in path [H1, H2]	UCA13: *follow other vehicles* provides a large longitudinal deceleration when another vehicle close behind [H3]		
		UCA14: *follow other vehicles* provides a large longitudinal deceleration on a low mu surface [H4]		
		UCA15: *avoid obstacles* provides a large longitudinal deceleration when another vehicle close behind [H3]		
		UCA16: *avoid obstacles* provides a large longitudinal deceleration on a low mu surface [H4]		

Table 12: UCAs for the HAS *Longitudinal Acceleration / Deceleration* control action

UCA1	*Maintain position in lane* does not provide target trajectory changes to maintain vehicle's central lane position [H1]
Functional Insufficiency / Target trajectory	(1a) the driver uses automation when it is unsafe to do so
	(1a, 1b) wrong safe target trajectory calculated (e.g. lane boundary mis-detection, HD map data error)
	(2b) vehicle movement does not correspond to target trajectory (e.g. low mu surface)
	(3b) driver assumes HAS is controlling the vehicle's lateral position (e.g. mode confusion)
Constraints	CC1: HAS shall continuously determine a safe target trajectory while the vehicle is in motion
Requirements	R1: The driver's handbook shall clearly describe HAS performance capabilities and limitations
	R2: The use of HAS shall be inhibited below 4^{o} C
	R3: Driver audio visual display shall clearly indicate when HAS is controlling vehicle lateral position
UCA2	*Avoid obstacles* does not provide target trajectory override when object in path [H2]
Functional Insufficiency / Target trajectory	(1a) the automation does not cancel when driver tries to override to avoid object
	(2a) driver sees object, but their control input is too small to override automation
	(3a) the driver is slow to detect object in path because they have become inattentive
	(3b) the driver sees an object in vehicle's path, but expects automation to mitigate
Constraints	CC2: HAS shall deploy diverse robust means to determine safe target trajectory
Requirements	R4: HAS shall determine safe target trajectory from multiple diverse data sources
	R5: HAS shall detect data loss used to calculate safe target trajectory within x ms and return control to driver
	R6: HAS shall cancel when a driver steering angle of x rad or a steering rate of y $\mathrm{rad\ s^{1}}$ is detected
UCA3	*Right-of-way decision* does not provide target trajectory override when lane change required [H2]
Functional Insufficiency / Target trajectory	(1a) the automation fails to cancel when driver tries to override to change lane
	(3b) the driver sees the lane is ending and believes the automation will complete a lane change manoeuvre
Constraints	CC3: The driver's ability to override the target trajectory while automation is active shall be assured
Requirements	Requirements R1 and R6 applies

Table 13: *Target trajectory* UCA1 - 3 loss scenarios, constraints and requirements

UCA4	*Maintain position in lane* provides target trajectory change when no path deviation is required [H1]
Functional Insufficiency / Target trajectory	(1a) automation incorrectly calculates safe target trajectory given vehicle's relative position in lane
	(1b) worn road markings means lane boundaries incorrectly determined
	(1b) HD map data has not been updated to reflect latest road geometry
Constraints	Constraint CC2 applies
Requirements	Requirements R4 and R5 apply
UCA5	*Maintain position in lane* provides target trajectory (centring) while *avoid obstacles* is also changing target trajectory to avoid object in path [H2]
Functional Insufficiency / Target trajectory	(1a) automation failure means target trajectory control does not cancel when driver attempts to override
	(3d) driver applies steering input to override automation, but stops when they feel an opposing force being applied by the automation
	(4b) driver applies only a small steering override (insufficient to cancel automation), which the automation immediately corrects
Constraints	CC4: Upon detection of the driver attempting to override the automation all HAS control actions shall cease within *x* ms
Requirements	Requirement R6 applies
	R7: All automation control actions shall cease within *x* ms of the driver applying a steering, braking or accelerating control action
UCA6	*Maintain position in lane* provides target trajectory (centring) while *right-of-way decision* is also changing target trajectory to complete a lane change manoeuvre [H2]
Functional Insufficiency / Target trajectory	(1a) automation failure means target trajectory control does not cancel when driver attempts to override
	(3d) and (4b) as for UCA5 above
Constraints	Constraint CC4 applies
Requirements	Requirement R7 applies

Table 14: *Target Trajectory* UCA4 - 6 loss scenarios, constraints and requirements

UCA7	*Avoid obstacles* provides target trajectory change while *maintain position in lane* is also changing target trajectory (i.e. an accumulative effect) [H4]
Functional Insufficiency / Target trajectory	(2a) accumulative effect of driver and automation target trajectory change exceeds traction limits for given conditions / manoeuvre
	(3b) driver is unaware of the target trajectory already being applied by the automation, so over compensates with their input
	(3b, 3d) the automation target trajectory modification coincides with the driver seeing an object in path, which the driver misinterprets as HAS taking avoiding action
Constraints	CC5: the vehicle's stability control system shall be capable of overriding driver and automation steering inputs in order to maintain vehicle stability
Requirements	R8: the stability control system shall be capable of mitigating the accumulative effect of a simultaneous HAS and driver target trajectory change
	R9: driver deactivation of vehicle stability control system shall be prevented while HAS is active
UCA8	*Right-of-way decision* provides target trajectory change while *maintain position in lane* is also changing target trajectory (i.e. an accumulative effect) [H4]
Functional Insufficiency / Target trajectory	(1a) accumulative effect of driver and automation target trajectory changes leads vehicle to exit own lane
	(2a) accumulative effect of driver and automation target trajectory change exceeds traction limits for given conditions / manoeuvre
	(3d) the greater than expected vehicle lateral response causes the driver to panic and overcorrect
Constraints	Constraint CC5 applies
Requirements	Requirements R8 and R9 apply

Table 15: *Target Trajectory* UCA7 - 8 loss scenarios, constraints and requirements

automation, meaning that the driver's action does not avoid the obstacle in the vehicle's path. As a mitigation a safety requirement (e.g. R6 Table 13) defining the magnitude of steering wheel input needed to cancel the automation could be written.

Considering the vehicle plant model and *scenarios related to the controlled process (2b)* in the context of the UCAs identifies further loss scenarios. Perhaps, the ambient temperature is sufficiently low for ice to form on the road surface. Consequently, the *target trajectory* control action required by the automation does not translate into the corresponding *complete vehicle movement* because of the low mu surface. Such a scenario could be avoided by inhibiting HAS use below a certain ambient temperature (e.g. 4^0C in the case of safety requirement R2 Table 13).

C.3.5 3 – *Event evaluation and construct maintenance*

For HAS all loss scenarios involving event evaluation and construct maintenance identified relate to the driver. With all UCAs relating to *target trajectory* including at least one loss scenario of this type.

The analysis of UCA1, UCA3 and UCA7 identify loss scenarios relating to driver confusion regarding the automation's current state or automation's capability. The loss scenario for UCA1 is a classic case of *mode confusion*. Here, the driver assumes that the automation is controlling the vehicle when it is not, which means the vehicle's correct position in lane is not maintained. For UCA3 which relates to the behavioural competency *right-of-way decision* a loss scenario occurs because the driver does not understand the capability of the automation. Seeing the lane ending ahead, the driver assumes that HAS includes lane change functionality. The analysis of UCA7 identifies two loss scenarios in which vehicle instability results. In the first loss scenario the hazard occurs because of an accumulative steering input effect; with both the automation and the driver making a steering input together. The second loss scenario is another example of a system capability misunderstanding. Here, the driver's awareness

of an object in the vehicle's path is coincident with the automation making a correct steering correction. Consequently the driver incorrectly assumes that the automation has 'seen' the collision risk and is steering the vehicle to avoid the obstacle.

The loss scenario identified for UCA2 is perhaps the result of a combination of effects. Firstly, the automation has failed to detect the presence of an object posing a collision risk in the vehicle's path. This situation is exacerbated by the driver's inattentiveness, which negatively impacts the time to *evaluate the event (3a)* and to avoid the hazard.

For UCAs 5, 6 and 8 scenarios where a *different or unexpected behaviour (3d)* leads to the hazard have been identified. In UCA5 and UCA6 the driver starts to turn the steering wheel because they have identified an imminent collision risk or the need to change lane. However, feeling the automation making an opposition steering input, the driver stops making their steering input. This results in the object in path not being avoided. The loss scenario for UCA8 represents the opposite effect. In this case the magnitude of the driver's steering correction is too large. Feeling the automation making a steering correction that is coincident with the automation's, the driver panics and makes an excessive steering correction.

C.3.6 *4 – Choice and execution of control actions*

Two loss scenarios are identified where the agents choice or execution of a control action leads to the hazard. These loss scenarios related to UCA5 and UCA6, where the driver is attempting to make a steering correction at the same time that the automation is also changing the vehicle's *target trajectory*. In both cases, the driver's input is not large enough to cause the automation to cancel. As a consequence the vehicle either fails to avoid the object in path (UCA5) or the vehicle remains in the lane that is soon to end (UCA6).

C.4 ANALYSIS COMPARISON

A surprising omission from NHTSA's report is no consideration of a vehicle instability hazard, particularly given that ALC has direct control of vehicle steering, and hence the vehicle yaw.

Both this analysis and the NHTSA report have identified the transfer of control between the human driver and the automation as potential hazard causes. The focus of the NHTSA analysis was the early or late cancellation of ALC and driver attentiveness, while this analysis has focused on manoeuvres where the driver has not been able to override the automation, or has suffered *mode confusion*. For example, not being able to steer to avoid an obstacle undetected by the automation. The assertion is that if this analysis had fully explored the *driver take back request* control action, rather than just focusing on the *Target Trajectory* control action, then it would have uncovered more hazards causes relating to the transfer of control between the driver and automation.

The comparison of the two analyses is hampered by differences in the level of abstraction considered. The NHTSA report focuses on the Item whereas this analysis has a vehicle level focus. Consequently, this analysis has uncovered hazard causes relating to driver override failing (during an emergency manoeuvre), driver and automation shared inputs combining (resulting in vehicle instability), and driver mode confusion. In contrast, the NHTSA report identifies interactions with other vehicle systems, which is at an abstraction level below that considered by this analysis.

A review of the appendices to [21] has been undertaken. Again, it was difficult to make a direct comparison between that work and the loss scenarios considered here, due to the differences in abstraction. The appendices [20] focused on aspects such as sensor failure and system interactions, which are the types of failure typically uncovered by an FMEA.

D

AUTOMATIC LANE KEEPING SYSTEM CASE STUDY

D.1 INTRODUCTION

This section demonstrates the use of the EVCM (see Chapter 8, page 112) with the *Shared Control* STPA method (see Chapter 9, page 123) using the Automatic Lane Keeping System (ALKS) vehicle feature as a worked example. This demonstrates how the EVCM is used with the *Shared Control* STPA method to identify hazard causes potentially resulting from shared control. This section concludes with the shared control safety case argument pattern (see Chapter 10 page 144) being instantiated for the ALKS example.

Following the steps shown in Figure 12 (Chapter 7, page 110), Section D.2 identifies the behavioural competencies applicable to ALKS. It then considers them in the context of the ODD and reflects on what behavioural competencies are necessary to maintain safety in the ODD. The behavioural competencies are then used to populate the EVCM, which forms the CSD for the later analysis. Section D.3.3 then uses the STPA guide words to identify the UCAs for ALKS before analysing those UCAs in the context of shared control to identified applicable loss scenarios. This identifies the safety requirements (system and controller level constraints) for ALKS. The loss scenarios identified in Section D.3.3 then inform the test strategy (see Section D.3.4). The chapter concludes with a summary.

D.2 BEHAVIOURAL COMPETENCY SELECTION

D.2.1 *ALKS behavioural competencies*

Regulation 157 has recently been published in relation to ALKS [180]. It specifies the requirements that an ALKS vehicle feature must achieve before an M1 class vehicle fitted with ALKS may receive type approved [179]. A qualitative analysis of Regulation 157's requirements, using the NVivio qualitative data analysis software identifies 12 behavioural competencies related to the DDT and 3 competencies related to the automation itself. The NVivo "sunbursts" (Figure 34) provides a pictorially representation of the behavioural competency hierarchy. In sunburst (a) *Avoid Obstacles* and *Transition of Control* are coloured in a darker shade, to represent that they have the most coding instances. Sunburst (b) then shows the 'spread' of Regulation 157 requirements. The behavioural competencies[1] together with their interpretation for ALKS, are described below:

- Strategic Level Competencies
  - COMPLY WITH RULES: The Regulation specifies that, while active, ALKS's DDT must comply with traffic rules in the country of operation.
- Manoeuvring Level Competencies
  - AVOID OBSTACLES: ALKS shall not cause a collision that is reasonably foreseeable or preventable. On detecting a collision risk the system must bring the vehicle to a stop within its lane. In this context collision risk includes a non-obstructed pedestrian. The forward detection range of ALKS must be 46 m in front of the vehicle and must include the left- and right-hand adjacent lanes. The Regulation defines the cut-in, cut-out and sudden deceleration by the preceding vehicle as scenarios that the automation must be able to cope

1 Appendix A gives a generic interpretation of each behavioural competency.

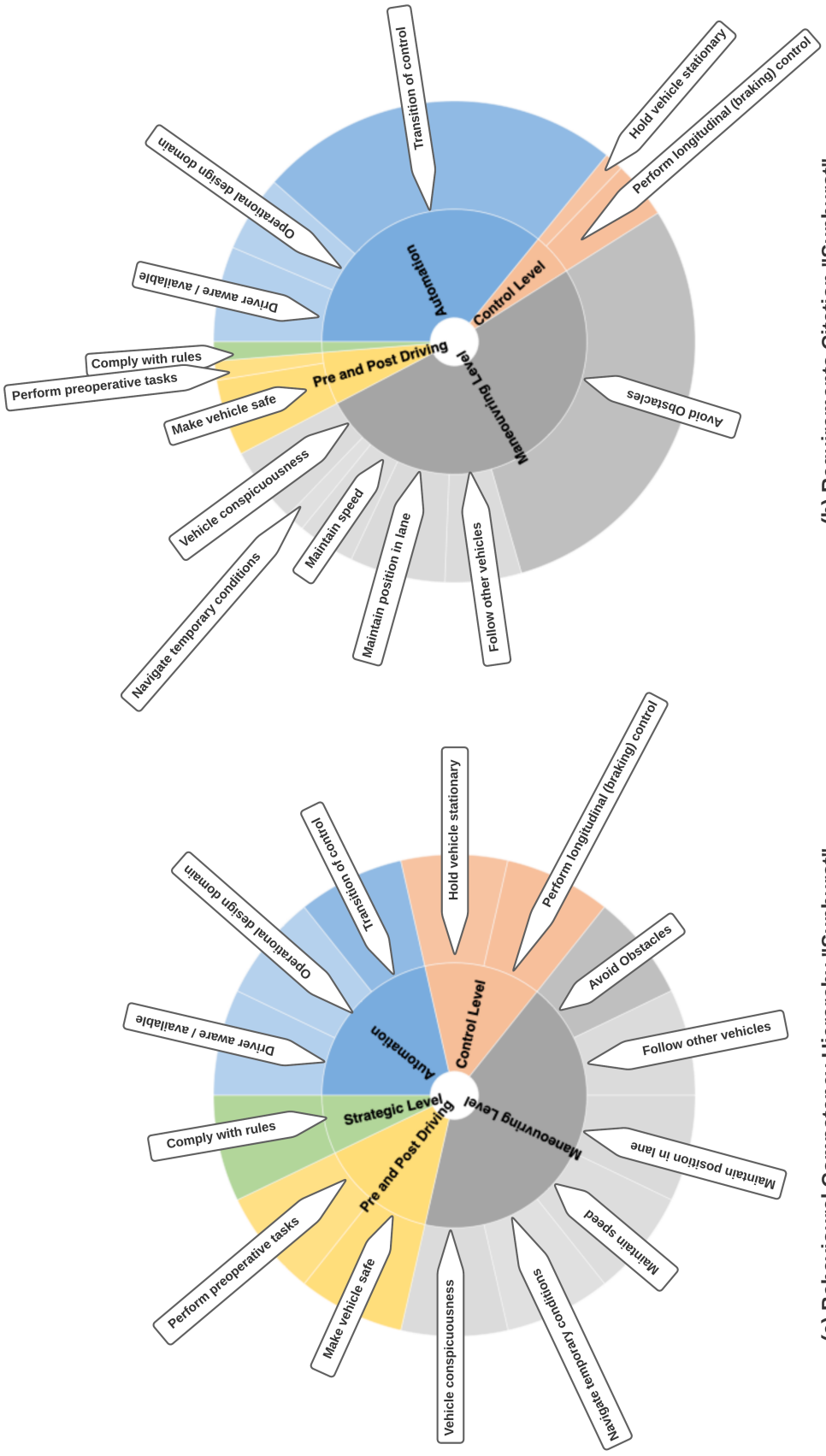


Figure 34: NVivo qualitative analysis of ALKS Regulation 157 requirements [180]

with. This competency is considered analogous to *avoid obstacles* behavioural competency described in Appendix A.

– EMERGENCY MANOEUVRE: In the context of ALKS this behavioural competency is closely related to *make vehicle safe* discussed below. With the exception of a situation where the Regulation considers a human driver would fail to avoid a collision (see *avoid obstacles* above) ALKS shall be capable of bringing the vehicle to a stop.

– FOLLOW OTHER VEHICLE: When following another vehicle ALKS must adjust the vehicle speed to maintain a minimum safe distance to the preceding vehicle. This includes managing vehicle speed from stationary and dealing with dynamic infringements of this rule; for example, when an adjacent vehicle cuts-in dramatically reducing the *'headway'* to the preceding vehicle. The minimum following distance shall never be less than 2 m.

– MAINTAIN POSITION IN LANE: While active ALKS must maintain the vehicle in a central position in lane, and must ensure that the vehicle does not cross any lane markings. The system must also take account of the relative position of adjacent road users, and be capable of modifying the vehicle's lateral position in lane or speed accordingly.

– MAINTAIN VEHICLE SPEED: The maximum vehicle speed for ALKS is 60 km/h. While active, ALKS must be capable of controlling vehicle speed. This includes adjusting the vehicle speed to take account of infrastructural and environmental conditions (e.g. narrow curves, inclement weather).

– NAVIGATE TEMPORARY CONDITIONS: The Regulation captures this behaviour with just one requirement! An *unplanned event* is defined as *"a situation which is unknown in advance, but assumed as very likely in happening, e.g. road construction, inclement weather, approaching emergency vehicle, missing lane marking, load falling from truck (collision) and which requires a transition demand."* When such an *unplanned event*

occurs the system is expected to immediately transition control back to the driver.

  - VEHICLE CONSPICUOUS: If ALKS has brought the vehicle to a standstill (e.g. following an evasive manoeuvre or Minimum Risk Manoeuvre (MRM)) then the vehicle's hazard warning lights must be activated to warn other road users.

- Control Level Competencies
  - PERFORM LONGITUDINAL (BRAKING) CONTROL: To support the requirement to not cause a collision, the system must be capable of decelerating the vehicle up to its full braking performance.
  - PERFORM LONGITUDINAL (ACCELERATION) CONTROL: The Regulation does not include any specific requirements relating to longitudinal acceleration. However, the competency is included here for completeness, as it would be the control level competency supporting both *follow other vehicles* and *maintain vehicle speed*.
  - PERFORM LATERAL CONTROL: Like *perform longitudinal (acceleration) control*, no specific requirements exist for this control level competency. Again it has been included for completeness, because it supports the manoeuvring level behavioural competency *maintain position in lane*.

- Pre and Post Driving Competencies
  - PERFORM PREOPERATIVE CHECKS: The system shall continuously carry out self-checks to detect failures and to confirm the system's performance. This includes confirming at start-up that an object can be detected at least 46 m in front of the vehicle.
  - MAKE VEHICLE SAFE: This competency can be thought of in the context of a Minimum Risk Manoeuvre (MRM) required by the Regulation. The Regulation requires that the vehicle is brought to a standstill, following an MRM, unless ALKS is deactivated (presumably by the driver) prior to the vehicle becoming stationary.

The ALKS regulatory requirements highlight three further non-DDT applicable competencies: *Assess driver availability*, *ODD detection* and *transition of control*:

- Automation use
  - ASSESS DRIVER AVAILABILITY: With the expectation being that ALKS will be unable to manage vehicle safety under all conditions, there is a requirement to continually monitor the driver's attentiveness; to ascertain their availability to resume manual control if required. This includes monitoring the driver's presence with seat belt fastened (to ascertain that the driver remains sat behind the controls), and monitoring the driver's gaze (to access their level of attentiveness).
  - ODD DETECTION: The Regulation requires that ALKS is used *"on roads where pedestrians and cyclists are prohibited and which, by design, are equipped with a physical separation that divides the traffic moving in opposite directions."* The Regulation also requires the implementation of strategies which detect and compensate for environmental conditions that could affect the system's detection range.
  - TRANSITION OF CONTROL: As one might expect, the Regulation places extensive requirements on the transition of control from the driver, to the automation and back to the driver again. If the system can no longer meet the requirements of the Regulation then ALKS is disabled. The system must be capable of recognising all situations where the transition of control back to the driver is needed, and afford the driver sufficient time / warning to safely resume control. If the driver fails to resume control then the MRM is triggered. The transition of control competency also includes managing driver overrides – via the brake or accelerator pedals, or via the steering hand wheel.

D.2.2 *Behavioural competencies for safety*

The thematic analysis approach, described in Section D.2.1, captures the behavioural competencies described by Regulation 157's requirements. Considering the behavioural competency taxonomy (see Appendix A, page 210) identifies further behavioural competencies not identified in the Regulation, but required to maintain vehicle safety in the typical ODD scenarios envisaged (e.g. operating on a UK 'Smart Motorway' in moderate to high traffic density). This analysis is captured in Table 16.

Some behavioural competencies are considered out-of-scope for ALKS. For example, *start the vehicle, route planning*, and *perform a u-turn*. Others are obviously in scope and under the sole responsibility[2] of ALKS. For example, *follow other vehicles, maintain position in lane* and *avoid obstacles*.

The responsibility split between the actor 'doing', 'monitoring' and being responsible for vehicle 'safety' becomes less clear for behavioural competencies like *right of way decision, navigate temporary conditions* and *respond to traffic conditions*. Clearly, the ability to determine whether the subject vehicle has right of way in a given scenario is in scope – particularly in the context of UK 'Smart Motorways' where a red 'X' displayed above the lane signifies a lane closure. However this review (see Section D.2.1) failed to identify any specific requirements relating to this competency. Therefore, the initial conclusion was that 'doing', 'monitoring' and 'safety' are all the human driver's responsibility and so should be shaded green. However, if we assume that the driver may undertake non-driving tasks (e.g. interact with an audio streaming service), while ALKS shall comply with the road traffic rules applicable to the operational region, then it is probably more reasonable to assume that 'doing' and 'monitoring' become shared responsibilities. While the responsibility for 'safety' remains with the human driver.

2 The Regulation does not explicitly state whether the driver is expected to undertake non-driving tasks while ALKS is active. However, given that there are requirements relating to the arbitration of ALKS screen messages, this analysis assumes that the driver will involve themselves in secondary tasks while ALKS is active – i.e. interfacing with other vehicle systems.

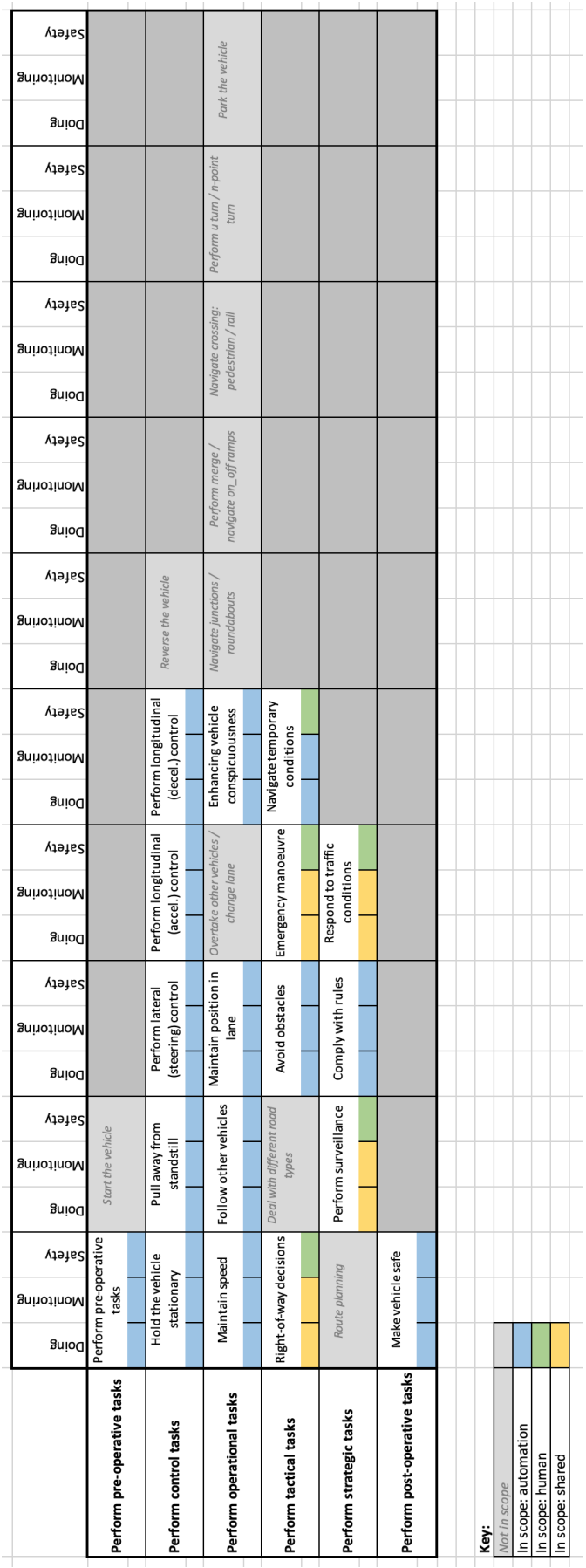


Table 16: ALKS actor responsibility behavioural competency comparison: who is doing, who is monitoring and who is responsible for vehicle safety?

The behavioural competency *emergency manoeuvre* has been treated similarly. The ALKS emergency manoeuvre behaviour defined by the Regulation (termed the MRM) relies on braking as a mitigation. If an emergency manoeuvre requires a significant steering input to mitigate the event (e.g. vehicle instability due to a low mu surface) ALKS may not have the authority over the vehicle controls to mitigate such a situation. It such circumstances the onus would be on the driver to understand potential situations when the system might fail to deal correctly with a situation. Consequently 'doing', 'monitoring' and 'safety' are designated 'shared', 'shared' and 'human' respectively.

Navigate temporary conditions has been designated as the automation being responsible for 'doing' and 'monitoring' and the human being responsible for 'safety'. This is because the Regulation contains requirements relating to the system's ability to detect *planned* and *unplanned events* within the ODD that require the transition of control, but with control returned to the driver in such cases the human would remain ultimately responsible.

Two further behavioural competencies were identified by the thematic analysis as being undertaken solely by the human driver – *perform surveillance* and *respond to traffic conditions*. However, if the assumption is made that the driver can remove their eyes from the road then the automation must contribute to these behavioural competencies in some way too. As stated above, the Regulation expects the automation to identify when conditions mean that DDT control needs to be returned to the driver. However, it also includes requirements highlighting the need to inform the driver (e.g. giving guidance in the Driver Handbook) regarding the capability / performance limitations of the ALKS feature. This need to correctly interpret user document and enable / disable the automation as appropriate, has lead these behavioural competencies to be designated as: 'doing' shared (shaded yellow) and 'monitoring' shared (shaded yellow), and 'safety' the responsibility of the human driver (shaded green).

Figure 35 depicts the interaction between the behavioural competencies discussed. As expected from an SAE Level 3 system, the correct operation of the *transition of control* behavioural competency is pivotal to ALKS safety. To focus

Stakeholders
The driver of the vehicle
Vehicle occupants
Other road users
Insurance companies
Vehicle manufacturer
Type approval authorities

Losses	
L1	Loss of life or injury to people
L2	Loss of or damage to the vehicle
L3	Loss of or damage to objects outside the vehicle
L4	Loss of consumer confidence or sales
L5	Financial loss
L6	Loss of liberty

Hazards		Related Losses
H1	Subject vehicle does not maintain a safe distance to the preceding vehicle	L1, L2, L3, L4, L5, L6
H2	Subject vehicle does not maintain a safe distance to adjacent vehicles	L1, L2, L3, L4, L5, L6
H3	Succeeding vehicle cannot maintain a safe distance due to subject vehicle's deceleration rate	L1, L2, L3, L4, L5, L6
H4	Subject vehicle fails to stop for / avoid an object in path	L1, L2, L3, L4, L5, L6
H5	Subject vehicle stays from own lane	L2, L3, L4, L5
H6	Subject vehicle loss of stability / control	L1, L2, L3, L4, L5, L6
H7	Subject vehicle fails to adhere to applicable road traffic laws	L5, L6

Table 17: ALKS applicable stakeholders, stakeholder losses, and vehicle level hazards

the attention on the manoeuvring level tasks and to avoid diagram clutter the "control level" behavioural competencies are not fully developed in the diagram.

D.3 THE ANALYSIS OF ALKS

This section progresses through the five steps of the *Shared Control* STPA method as depicted in Figure 12 (Chapter 7, page 110).

D.3.1 *Define the purpose of the analysis*

Step 1 of the STPA method defines the purpose of the analysis. Using expert judgement and considering ALKS use in the context of a multi-lane highway[3], applicable stakeholders, stakeholder losses and vehicle level hazards are identified. These appear in Table 17.

From the hazard list the system constraints are derived (Table 18). These are somewhat analogous to the safety goals typically derived for automotive systems, although defined at the vehicle rather than at the Item level.

3 That is, a multi-lane dual carriageway having a physical barrier separating the traffic moving in opposite directions.

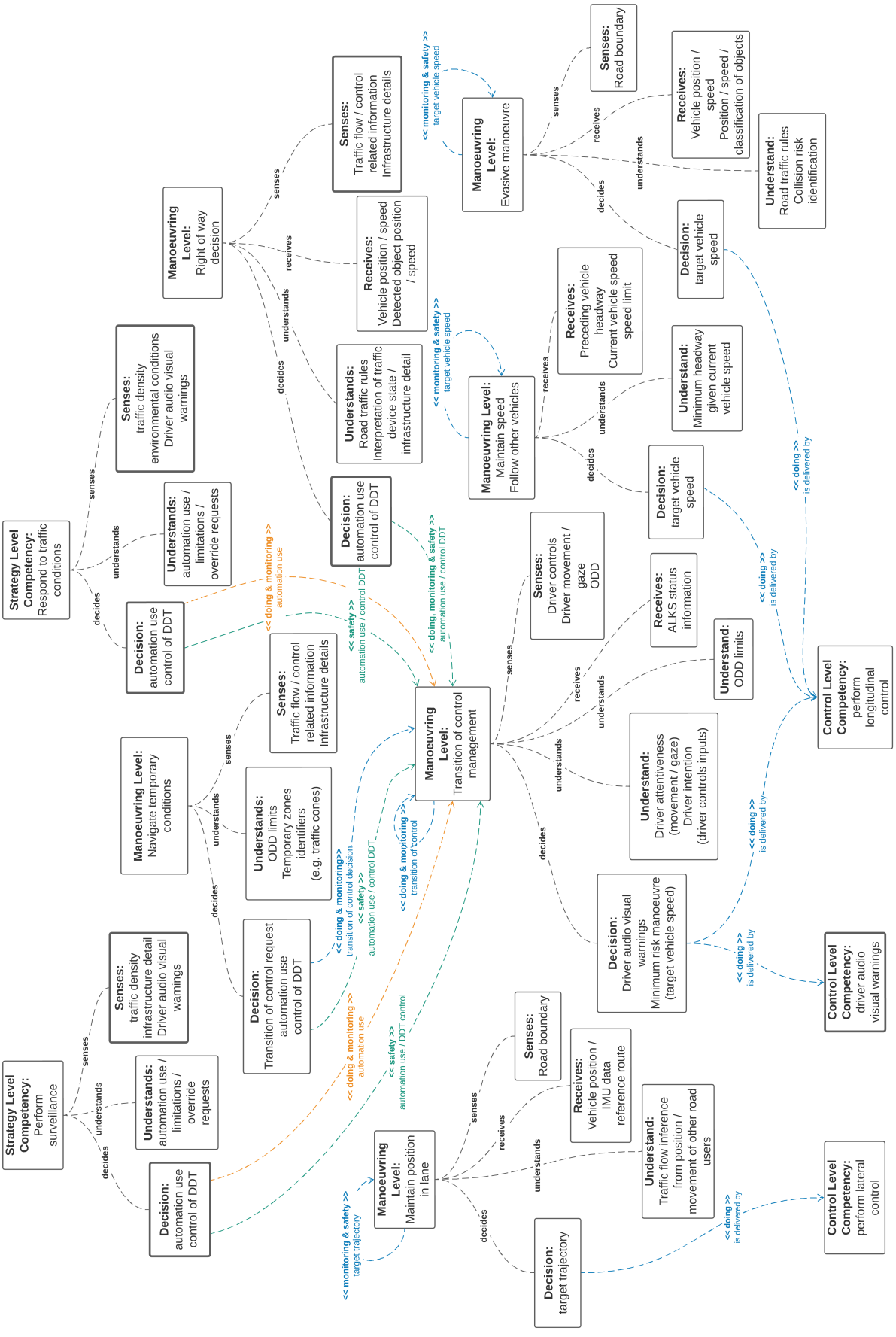


Figure 35: ALKS behavioural competency interaction diagram

Safety Constraint	Description	Related Hazard
SC1	The subject vehicle shall maintain a safe headway to the preceding vehicle of $> x$ s.	H1
SC2	The subject vehicle shall maintain a safe distance to adjacent vehicle of $> y$ m while the vehicle is in motion.	H2, H5
SC3	The subject vehicle shall avoid non-emergency brake applications of $> - z$ ms^{-2}.	H3
SC4	The subject vehicle shall avoid collisions with objects in path.	H4
SC5	The subject vehicle shall maintain a central lane position and avoid path deviations of $>$ 0.3 m.	H2, H5
SC6	The subject vehicle shall avoid a change in yaw rate of $> 2.5^{o}\,s^{-1}$.	H6
SC7	The subject vehicle shall adhere to applicable road traffic laws.	H7

Table 18: ALKS derived system constraints (Safety Goals)

D.3.2 *Control structure diagram*

The CSD for ALKS (see Figure 36) is developed by populating the EVCM with the identified behavioural competencies (see Section D.2.2), control actions and pertinent feedback.

D.3.3 *Unsafe control actions, functional insufficiencies, triggering conditions and safety constraints*

As identified by the behavioural competency interaction diagram (see Figure 35, page 282), it is the *strategy* and *manoeuvring* level competencies where interactions of significance take place between the automation and the human driver.

The below tables show the UCAs, functional insufficiencies, triggering conditions and safety constraints for each control action identified in Section D.3.2. The UCAs are identified by applying the guide words at the top of each

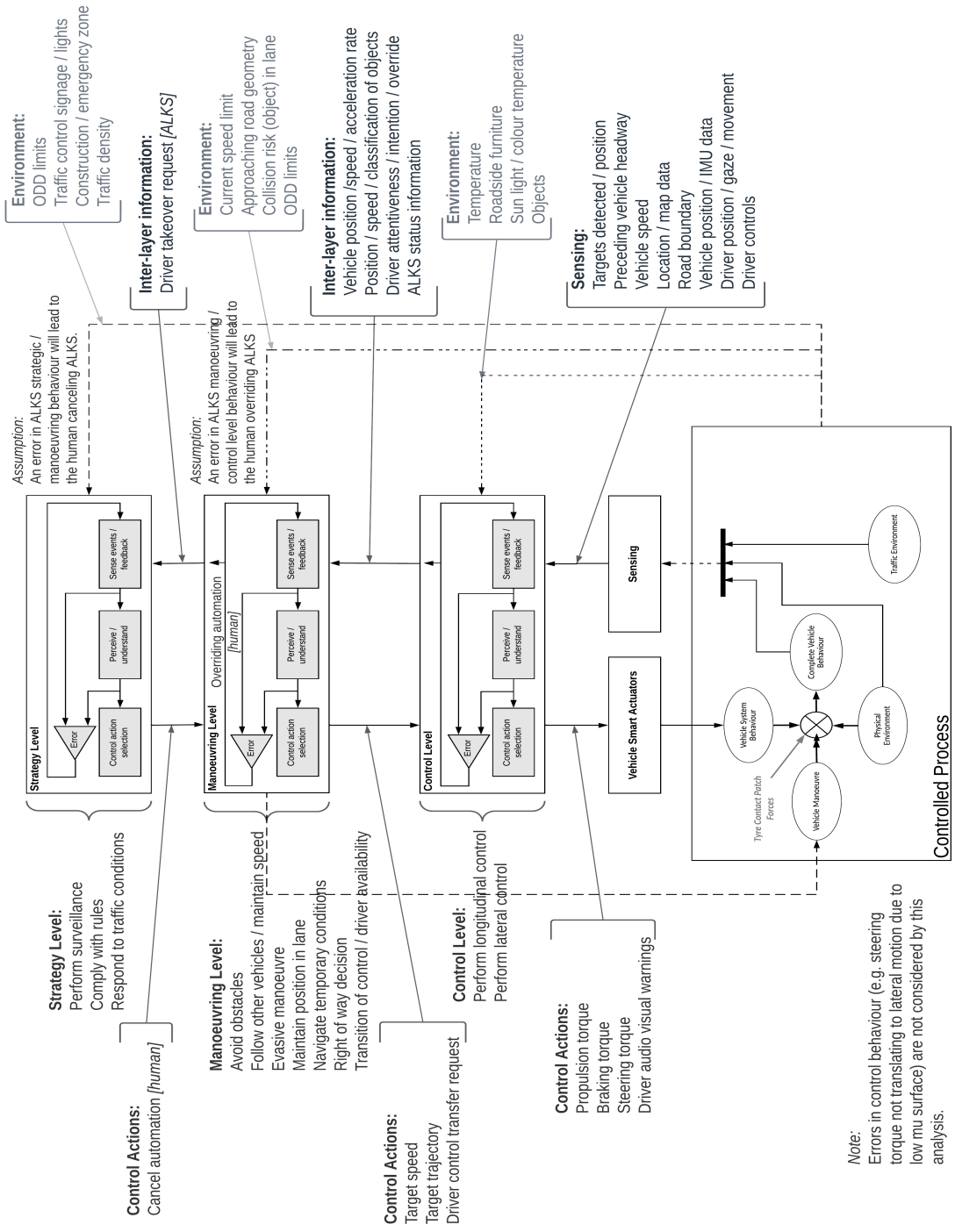


Figure 36: The CSD for ALKS

column. Loss scenarios are then identified from the UCAs. From the loss scenarios the potential functional insufficiencies and triggering conditions [165] are identified. The letters and numbers in brackets (e.g. "1a", "4b") identify the loss scenario type (see Figure 17, page 137) that uncovered the functional insufficiency or triggering condition. Finally, controller constraints are written to mitigate the UCAs. These safety requirements are then further developed and refined to mitigate the functional insufficiencies and triggering conditions identified. This analysis is captured in Tables 19 to 23 below.

D.3.4 *Test scenarios*

Tables 24 and 25 are included for completeness. They contain the loss scenarios generated for each UCA, used to identify the functional insufficiencies and triggering conditions discussed in Section D.3.3 above. The expectation is that simulation will be used to verify ALKS behaviour, and this loss scenario catalogue will inform the scenario based simulation testing undertaken.

D.3.5 *Safety case argument*

The final process step (i.e. step 5 in Figure 12, page 110) instantiates the safety case argument patterns. This follows the safety case argument structure first introduced in Chapter 10. It should be noted that this is not the complete safety case argument for ALKS, but solely the argument for shared control.

The shared control safety case argument (see Figure 37) starts with goal *G1: Vehicle accidents caused by the ALKS Feature's shared control are avoided*. This goal is supported by three further sub-goals. Goal G2 claims that the *Shared Control* STPA method finds all UCAs and loss scenarios caused by ALKS's shared control. Goal G3 claims that the safety requirements for ALKS mitigate all UCAs and loss scenarios identified. Finally, goal G4 makes the claim about the safety of ALKS's shared control, given that the feature has been tested in those loss scenarios known to be hazardous.

Control Action	Unsafe Control Actions			
	Not providing causes hazard	*Providing causes hazard*	*Provided but timing wrong*	*Provided but duration wrong*
Target trajectory	UCA1: ALKS does not provide target trajectory changes to maintain vehicle's central position in lane [H2, H4]	UCA2: ALKS provides a target trajectory change greater than the safe trajectory requires (e.g., straight or large radii curve) [H2]	UCA3: ALKS provides a target trajectory change after the vehicle has left the ODD and DDT control is returned to the driver (ALKS cancel) [H2]	UCA4: ALKS stops target trajectory change while the required safe trajectory is still changing (e.g., mid corner) [H2]
	Functional Insufficiency / Triggering Condition			
	(1a) the ALKS algorithm fails to determine the correct safe path from map data and video data	(1b) incorrect lane boundary information received from perception system	(1a) failure in the ALKS controllers overwrites manual target trajectory	(1a) failure in ALKS controller
	(1b) the ALKS perception system loses the position of the current lane boundary or misinterprets where the lane boundary actually is.	(4b) being concerned about a potential collision risk, the driver reacts with a steering input too large for the current speed		(2b) error in perception system or map data causes incorrect trajectory determination
	Constraints			
	CC1: ALKS shall continuously determine a safe target trajectory while the vehicle is in motion	CC2: ALKS shall constrain the target trajectory rate of change based on current operating conditions (e.g., vehicle speed, road curvature) so as not to exceed vehicle handling characteristics	CC3: ALKS modification of the target trajectory shall be prevented while the vehicle is operating in manual driving mode (ALKS disabled)	Constraint CC1 applies.
	Requirements			
	R1: ALKS shall be capable of detecting errors in map data received	R4: ALKS shall determine the maximum rate of change of target trajectory for operating conditions.	R6: the integrity of target trajectory data in memory shall be assured.	R8: the integrity of ALKS control computation shall be assured.
	R2: ALKS shall be capable of detecting errors in video data received.	R5: ALKS shall filter the target trajectory rate of change $\leqslant$ the safe rate of change.	R7: ALKS control of steering actuation shall be inhibited while vehicle is being operated manually.	Requirement R1: applies.
	R3: ALKS shall check the plausibility of map data in relation to the video images received			

Table 19: ALKS *Target Trajectory* UCAs, loss scenarios, constraints and requirements (i)

Control Action	Unsafe Control Actions			
	Not providing causes hazard	*Providing causes hazard*	*Provided but timing wrong*	*Provided but duration wrong*
Target trajectory	UCA5: ALKS does not provide a target trajectory change to modify lane position to maintain safe distance to adjacent object (e.g., wide load, narrow lanes in construction zone) [H2, H4]	UCA6: ALKS provides an excessive target trajectory change when DDT control returns to the driver [H2, H5]		UCA7: ALKS stops controlling target trajectory before the driver is ready to undertake the DDT [H2, H5]
	Functional Insufficiency / Triggering Condition			
	(1a) ALKS algorithm doesn't modify vehicle's position in lane when object protrudes into lane (e.g., stranded vehicle partially in live lane).	(3a) with driver out of the loop for long period, their 1st steering input is large. Perhaps misjudging the current vehicle speed too.		(1b) ALKS driver monitoring system suggests that the driver is attentive when they are not.
	(1b) 'blind spot' in perception data means ALKS not always capable of detecting vehicles encroaching the lane.			(3a) system doesn't give the driver sufficient time to become fully situationally aware.
				(4c) ALKS transfers control to the driver when difficult to regain situational awareness (e.g., high traffic density) or when the driver is required to act straight away (e.g., corner approaching).
	Constraints			
	CC4: ALKS shall be capable of detecting objects encroaching the vehicle?s path and modify the target trajectory / speed accordingly.	CC5: When a driver steering override is first detected ALKS shall limit trajectory rate of change (based on vehicle speed) to minimise excessive lateral acceleration / vehicle instability.		CC6: ALKS shall continue controlling target trajectory, maintaining a safe lane position until driver availability has been confirmed for x s AND request is acknowledged by the driver.
	Requirements			
	R9: ALKS perception algorithm shall be capable of detecting objects that encroach the vehicle's path and mitigate accordingly.	R4: ALKS shall determine the maximum rate of change of target trajectory for operating conditions		R11: Analysis and testing shall demonstrate the robustness of driver attentiveness determination – including reasonably foreseeable misuse
	R10: The capability of ALKS perception shall be demonstrated, particularly in the areas immediately adjacent to the vehicle	R5: ALKS shall filter the target trajectory rate of change $\leqslant$ the safe rate of change		R12: ALKS transition of control shall provide both audible and visual warnings that afford the driver a minimum of x s to become fully situationally aware before relinquishing DDT control.

Table 20: ALKS *Target trajectory* UCAs, loss scenarios, constraints and requirements (ii)

Control Action	Unsafe Control Actions			
	Not providing causes hazard	*Providing causes hazard*	*Provided but timing wrong*	*Provided but duration wrong*
Target speed	UCA08: ALKS does not provide target speed reduction when vehicle enters a lower speed limit zone [H6]	UCA9: ALKS provides excessive target speed reduction when succeeding vehicle is close behind [H3]		UCA10: ALKS stops reducing target speed before vehicle speed reduced sufficiently to maintain safe distance to preceding vehicle / object in path [H1, H4, H6]
	Functional Insufficiency / Triggering Condition			
	(3c) teaching data did not include 'smart motorway' signage	(1b) camera system incorrectly classifies 2D image as a collision risk		(1a) sampling rate of target speed / position not sufficient to detect rapid changes
	(3b) dirt / graffiti covering sign	(1b) road furniture / adjacent vehicle causes radar reflection in vehicle's path		(3a) traffic congestion causes preceding vehicle speed to isolate, minimising time available to detect speed changes
	(1b) sign back lit by strong sunlight	(4c) not expecting the preceding vehicle to brake suddenly the driver of the following car is slow to react		(4a) driver accidentally hits the accelerator pedal while fidgeting in seat.
	Constraints			
	CC7: ALKS shall be capable of detecting speed limit changes and modify the target speed accordingly	CC8: When no forward collision risk exists ALKS shall control the target speed rate of change such that vehicle deceleration does not exceed x ms^{-2}		CC9: ALKS shall only increase target speed once the headway to the vehicle / object ahead has been increasing for > 1 s
	Requirements			
	R13: ALKS training data shall include electronic signage as used on 'smart motorways'	R15: ALKS perception system shall use data fusion techniques, to improve robustness to 2D image / reflection false positives		R17: ALKS target detection / headway determination shall be robust to rapidly oscillating traffic flow
	R14: ALKS training data shall include signage affected by dirt / graffiti / lighting	R16: ALKS shall modify vehicle deceleration rate to optimise time available to preceding vehicle to keep clear		R17: Driver override algorithm shall filter driver accel. pedal input to minimise accidental system overrides

Table 21: ALKS *Target Speed* UCAs, loss scenarios, constraints and requirements (i)

Control Action	Unsafe Control Actions			
	Not providing causes hazard	*Providing causes hazard*	*Provided but timing wrong*	*Provided but duration wrong*
Target speed	UCA11: ALKS does not provide target speed reduction needed to maintain safe headway to preceding vehicle	UCA12: ALKS provides excessive target speed reduction when driver resumes manual control (brake pedal override, ALKS cancel)		UCA13: ALKS stops providing target speed reductions before the driver is ready to undertake the DDT (e.g., transition of control request at end of ODD)
	Functional Insufficiency / Triggering Condition			
	(1b) Incorrect vehicle speed information received	(1a) ALKS cancel causes sudden removal of longitudinal DDT control		(1a) ALKS transfer of control algorithm ignores external factors (e.g., currently braking in traffic) when requesting driver resumes DDT control
	(1b) Preceding vehicle misclassified in an adjacent lane	(1a) ALKS cancel causes sudden removal of longitudinal DDT control, exacerbated by vehicle being driven on incline		(3d) realising that the vehicle was braking, the driver rapidly tries to press brake, accidentally hitting the accelerator instead
		(3b) driver not pressing accelerator pedal sufficiently to maintain vehicle speed when ALKS cancels		(4c) being out of loop, the driver takes time to realise how much brake pedal force is needed to achieve required deceleration rate.
	Constraints			
	CC10: ALKS shall control the target speed such that the headway to the preceding vehicle remains greater than 2 m	CC11: When a driver brake override is first detected ALKS shall limit the target speed rate of change decrease for x ms to minimise harsh brake application and potential vehicle instability		CC12: ALKS shall continue controlling target speed to maintain the headway to preceding vehicles until driver availability has been confirmed for x s AND the transition of control request is acknowledged by the driver
	Requirements			
	R18: The integrity of vehicle speed signal shall be assured	R20: Vehicle simulation of typical scenarios shall inform the appropriate ALKS cancel deceleration rate		R21: ALKS transfer of control duration shall be increased when other road users are detected ahead or in adjacent lanes
	R19: ALKS perception system shall be robust to radar reflections caused by roadside furniture / structures			R22: Haptic feedback shall be used to help the driver understand the level of control being applied by ALKS prior to DDT transfer

Table 22: ALKS *Target Speed* UCAs, loss scenarios, constraints and requirements (ii)

Control Action	Unsafe Control Actions			
	Not providing causes hazard	*Providing causes hazard*	*Provided but timing wrong*	*Provided but duration wrong*
Driver transfer of control	UCA14: ALKS does not provide driver transfer request when environmental conditions limit sensor performance	UCA15: ALKS provides a driver transfer of control request that is not actioned by the driver causing MRM to be invoked	UCA16: ALKS provides driver transfer request too late for the driver to resume DDT control when required (e.g., before vehicle leaves ODD)	UCA17: ALKS stops issuing a driver transfer of control request before driver is able to recognise the request, causing MRM to be invoked.
	Functional Insufficiency / Triggering Condition			
	(1a) ALKS controller does not include the functionality to 'read' smart motorway signs	(1b) the driver is correctly positioned and ready to resume control. However, their presence / alertness is not detected by the system.	(3a) the driver has been out of the loop. Regaining situational awareness is hampered by the poor light conditions / lack of visibility	(1b) driver awareness feedback suggests that the driver is ready to resume DDT control
	(1b) the map data received has not been updated to reflect the temporary closure	(2b) the driver sees the transfer of control messages but doesn't know this has to be acknowledged	(4c) the traffic is busy / vehicles merging from adjacent lanes increases driver's workload while they regain situational awareness	(3b) driver unclear when the transition of control actually happens. I.e., they are unaware they have full DDT control
		(2b) environmental factors (bright sunlight) stops the driver seeing the transfer of control request		(4b) environmental conditions (e.g., low light, high traffic density) make it difficult for the driver to gain situational awareness
	Constraints			
	CC13: ALKS shall be capable of detecting lane ends and transfer DDT control to the driver x m before the lane end	CC14: All ALKS driver warnings shall include an audio AND visual component which the driver can identifying under all conditions	CC15: ALKS shall maintain safety by controlling DDT for x s after the transfer of control request is first issued	CC12 applies
	Requirements			
	R13: ALKS training data shall include electronic signage as used on 'smart motorways'	R25: ALKS related messages shall clearly communicate the next action that the driver must take	R21: ALKS transfer of control duration shall be increased when other road users are detected ahead or in adjacent lanes	R11: Analysis and testing shall demonstrate the robustness of driver attentiveness determination – including reasonably foreseeable misuse
	R23: ALKS map data shall be updated every x hours to include temporary conditions	R26: ALKS driver warnings shall be robust to all environmental factors (audible and visual) that could affect driver identification		Requirement R25 applies
	R24: Map data shall be deemed invalid if older than y hours old			R27: ALKS transfer of control duration shall be increased to account for environmental conditions detected

Table 23: ALKS *Transfer of Control* UCAs, loss scenarios, constraints and requirements

UCA	UCA Description	Loss Scenario
UCA1	ALKS does not provide target trajectory changes to maintain vehicle's central position in lane	ALKS is active and controlling the vehicle's target trajectory. However, the ego vehicle strays from a central lane position and either hits an adjacent vehicle [H2] or stationary object [H4]
UCA2	ALKS provides a target trajectory change greater than the safe trajectory requires (e.g., straight or large radii curve)	ALKS is active and is controlling the vehicle's target trajectory. The target trajectory selected is too large for the current operating situation causing the ego vehicle to cross its lane boundary.
UCA3	ALKS provides a target trajectory change after the vehicle has left the ODD and DDT control is returned to the driver (ALKS cancel).	The vehicle is being driven manually when the target trajectory changes causing the ego vehicle to exit its lane [H2].
UCA4	ALKS stops target trajectory change while the required safe trajectory is still changing (e.g., mid corner) [H2].	ALKS is active and controlling target trajectory. The road curvature is still changing (e.g., vehicle mid corner) but the target trajectory stops changing causing the ego vehicle to leave its lane.
UCA5	ALKS does not provide a target trajectory change to modify lane position to maintain safe distance to adjacent object (e.g., wide load, narrow lanes in construction zone)	ALKS is active and controlling the vehicle's target trajectory. The ego vehicle encounters an obstruction encroaching the lane. Because the system does not modify the target trajectory / speed accordingly the ego vehicle hits the obstruction [H4].
UCA6	ALKS provides an excessive target trajectory change when DDT control returns to the driver.	The driver is ready and ALKS is transferring control back to the driver. When steering control returns to the driver their steering input is too great and the ego vehicle exits its lane [H2] or loses stability [H5].
UCA7	ALKS stops controlling target trajectory before the driver is ready to undertake the DDT.	ALKS has request that the driver resume DDT control. Transfer of control completes and ALKS stops controlling target trajectory. However, because the driver fails to make any steering inputs the ego vehicle leaves its lane [H2].
UCA8	ALKS does not provide target speed reduction when vehicle enters a lower speed limit zone.	ALKS is active and the ego vehicle enters a new speed restriction zone. The system does not perceive this change. Consequently, the target speed is not reduced causing the ego vehicle to break the new speed limit [H6].
UCA9	ALKS provides excessive target speed reduction when succeeding vehicle is close behind.	The ALKS perception system classifies an object as an imminent collision risk, so carries out an emergency stop. Not expecting the vehicle ahead to suddenly stop, the succeeding vehicle hits the ego vehicle [H3].
UCA10	ALKS stops reducing target speed before vehicle speed reduced sufficiently to maintain safe distance to preceding vehicle / object in path.	The vehicle is being driven with ALKS active in congested motorway traffic. Preceding vehicles in the queue are changing their speed up and down rapidly (oscillatory behaviour) [H1]. ALKS detects an object in path so begins modifying target speed to slow the vehicle. With the obstacle still in path ALKS begins increasing target speed once more causing the vehicle to hit the obstacle [H4].
UCA11	ALKS does not provide target speed reduction needed to maintain safe headway to preceding vehicle.	ALKS is active. The ego vehicle is approaching a preceding vehicle in the same lane but fails to modify the target speed to maintain a safe minimum headway. If the preceding vehicle had to brake hard the ego vehicle would fail to stop [H1].

Table 24: Catalogue of loss scenarios pertinent to ALKS use

UCA	UCA Description	Loss Scenario
UCA12	ALKS provides excessive target speed reduction when driver resumes manual control (brake pedal override, ALKS cancel).	The vehicle is being driven with ALKS active and another vehicle is in close proximity behind. The driver cancels ALKS which results in a sudden deceleration which the succeeding vehicle fails to mitigate for [H3]. The driver cancels ALKS. The resulting sudden deceleration results in a vehicle instability which the driver fails to control [H5].
UCA13	ALKS stops providing target speed reductions before the driver is ready to undertake the DDT (e.g., transition of control request at end of ODD).	ALKS has requested that the driver resume DDT control. ALKS is slowing the vehicle (because of preceding vehicle / stationary object in path) as the transition of control is occurring. When the driver takes over the required vehicle deceleration is not continued, and the vehicle hits the preceding vehicle [H1] or object [H4].
UCA14	ALKS does not provide driver transfer request when a manual lane change is required (e.g., temporary lane closure).	ALKS is active and controlling the DDT. The lane ahead is closed (e.g., red X due to breakdown, police closure due to accident). However, ALKS does not issue a transfer of control request before the obstruction is reached [H4, H6].
UCA15	ALKS provides a driver transfer of control request that is not actioned by the driver causing MRM to be invoked.	ALKS is active and a situation has arisen (e.g., approaching end of ODD) which requires the driver to resume DDT control. ALKS has issued a transfer of control request, but because this has not been actioned ALKS stops the vehicle in the live lane (i.e., MRM). The vehicle is at risk of being hit from behind [H3].
UCA16	ALKS provides driver transfer request too late for the driver to resume DDT control when required (e.g., before vehicle leaves ODD).	ALKS is active and a situation arises requiring the driver to resume DDT control (e.g., fog reducing performance of perception system). The transfer of control request is issued, but driver feels rushed into resuming control, resulting in them making some large control inputs.
UCA17	ALKS stops controlling the DDT before the driver is ready to begin DDT control.	ALKS is active and has issued a transfer of control request. Before the driver is fully ready to resume DDT control ALKS stops controlling the DDT with undetermined consequences.

Table 25: Catalogue of loss scenarios pertinent to ALKS use (cont.)

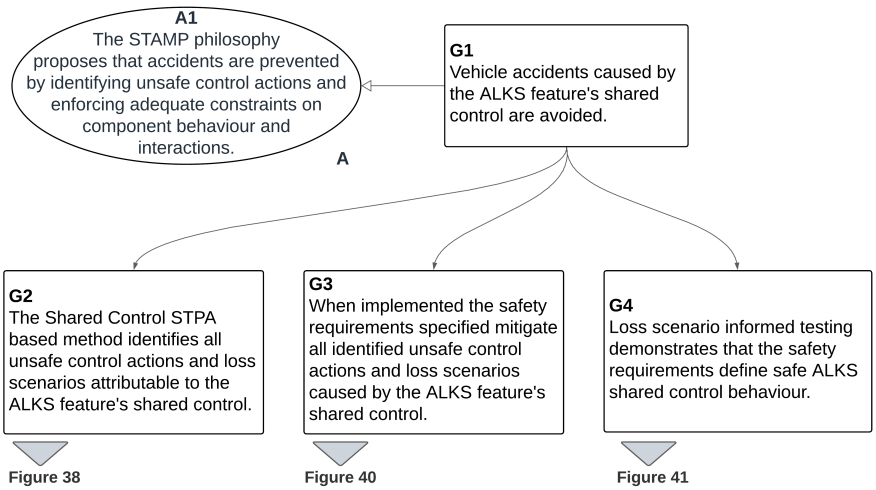


Figure 37: G1: ALKS shared control hazards are avoided

Figure 38 shows the instantiated safety argument structure for *G2: The Shared Control STPA based method identifies all UCAs and loss scenarios attributable to the ALKS Feature's shared control*. The evidence supporting this argument structure is drawn from Section D.2. Where a review of Regulation 157, in the context of the ODD, identified both the behavioural competencies delivered by ALKS and those needed to maintain the safety of ALKS (see Table 16, page 279). Representing the interaction between all relevant behavioural competencies (see Figure 35) then identified the shared control actions as: *Target Speed, Target Trajectory* and *Transfer of Control*. From this the CSD for ALKS was drawn, which provides the evidence for *G2.2: The EVCM instantiated with the shared control actions creates a CSD at an appropriate level of abstraction to facilitate the identification of shared control hazard causes*.

Systematically enacting the *Shared Control* STPA method generates the evidence supporting *G2.3* (see Figure 39). The information in Table 17 is typical of that defined by the *'Classic'* STPA method. The UCAs identified in Tables 19 to 23 are identified using the STPA guide words. Considering those guide words in the context of each relevant behavioural competency, uncovers further UCAs. The list of loss scenarios, identified by considering the identified UCAs in the context of both *'Classic'* (context *C6* in Figure 39) and *Shared Control* (context *C7* in Figure 39) loss scenario types, provides the evidence for goal *G2.3.3*.

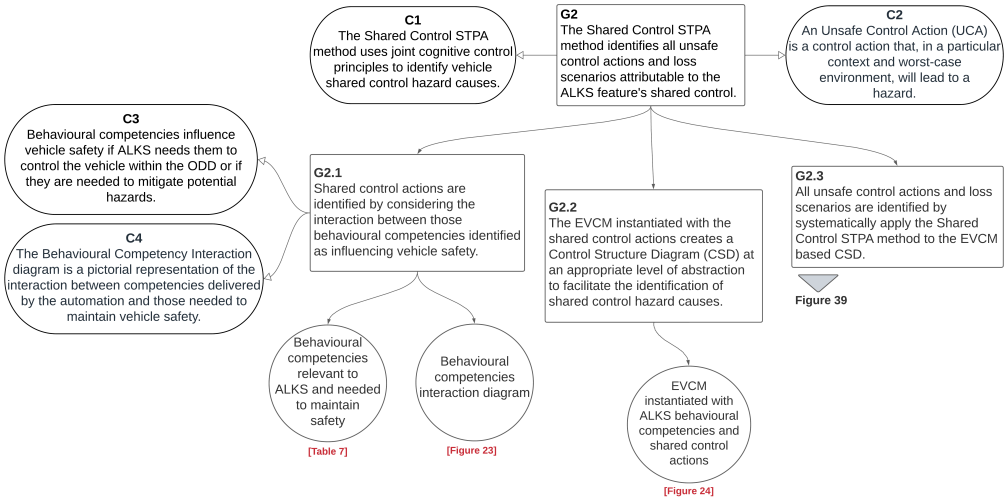


Figure 38: G2: All loss scenarios due to ALKS shared control are identified

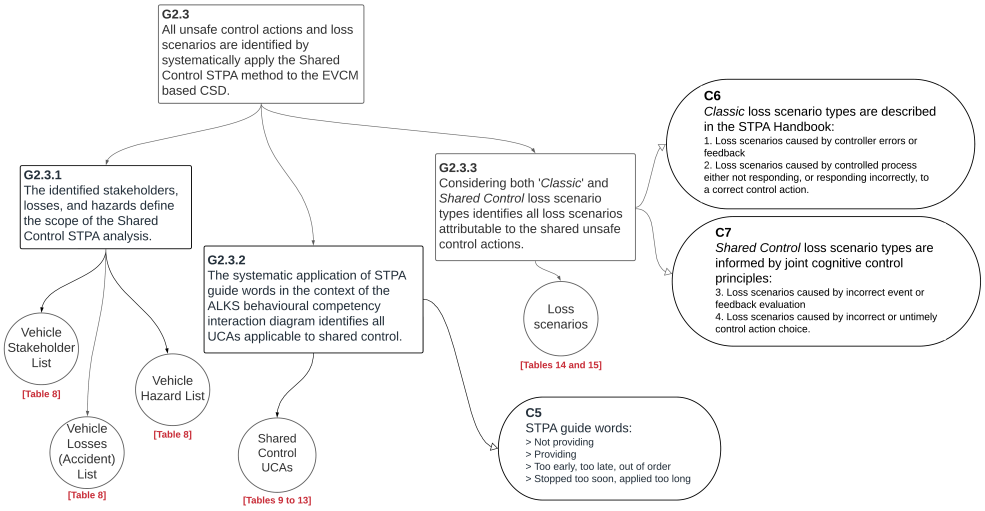


Figure 39: G2.3: Shared control STPA method applied systematically

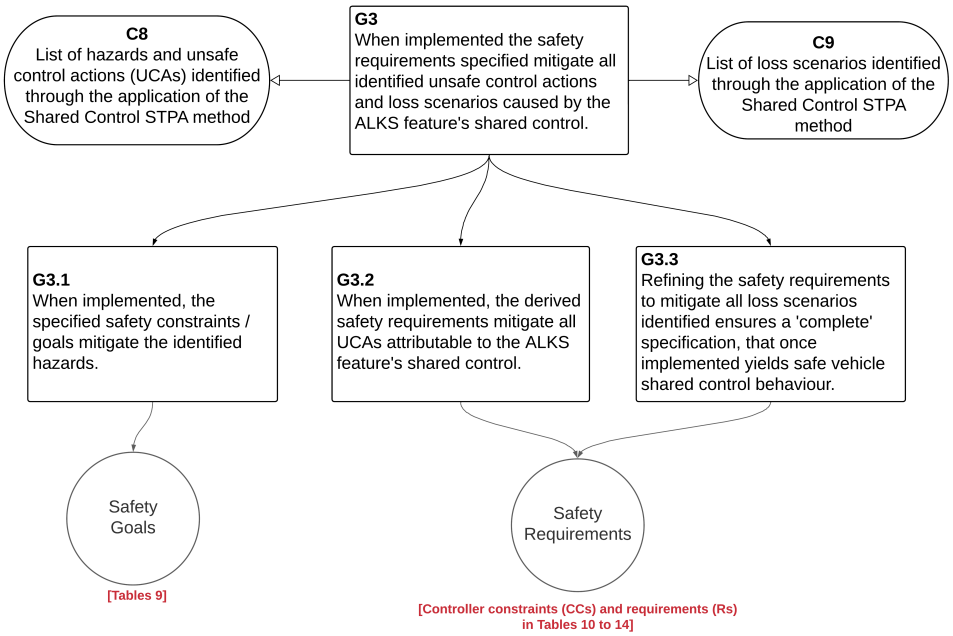


Figure 40: G3: ALKS safety requirements defined

Goal G3 (Figure 39) then relates to the safety requirements developed to maintain ALKS shared control safety. High level safety requirements (referred to by STPA as "safety constraints") are derived to mitigate the identified hazards. Further safety requirements (referred to by STPA as "controller constraints") are then developed to mitigate the UCAs identified. These are further refined by considering the loss scenarios identified. The safety goals and derived safety requirements form the evidence for goals: G3.1, G3.2 and G3.3 (Figure 40).

The final goal in the safety argument relates to testing – *G4: Loss scenario informed testing demonstrates that the ALKS feature exhibits safe shared control behaviour* (Figure 41). Not developed in this example, the expectation is that the loss scenario catalogue (i.e. Tables 24 and 25) will inform the scenario based simulation testing that would form a part of ALKS verification. Thus providing another dimension to the system integration and testing strategy, which is typically a requirements led strategy.

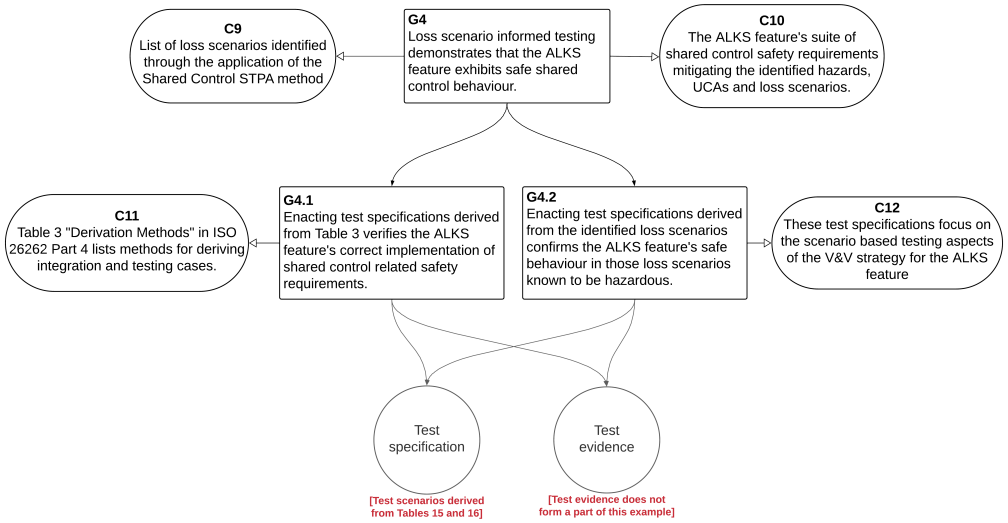


Figure 41: G4: ALKS shared control is demonstrably safe

D.4 SUMMARY

This section has used the ALKS vehicle feature (as specified by Regulation 157) [180] to demonstrate how the EVCM and the *Shared Control* STPA method can be used to analyse a feature whose correct shared control is important to safety.

Requirements relating to the transfer of control form a major part of the Regulation (see Figure 34). The review of the Regulation identified behavioural competencies at all three control hierarchy levels (i.e. strategic, manoeuvring and control). The behavioural competency interaction diagram (Figure 35 page 282) identifying that shared control occurred predominately at the manoeuvring level. This example does not show the STPA method being applied to the control level behavioural competencies. This could be done, and the expectation is that *'Classic'* STPA would uncover all loss scenario caused by the control level. That said, it is envisaged that the EVCM would prove a useful 'tool' in identifying external factors (e.g. environmental conditions) that potentially influence low-level closed loop control feedback paths.

The analysis uncovered many UCAs and loss scenarios that couldn't be claimed as "identifying anything *new*". For example, the importance of using only up-

to-date map data or making the driver warnings and displays robust to various ambient conditions. However, there were other situations that the author had not considered previously. For example, how traffic density might impact the time taken for the driver to regain situational awareness. Thus, requiring the system to potentially alter the point at which a transfer of control request is issued, based on traffic density. Also, identifying that haptic feedback could be deployed to help the driver regain DDT control more quickly.

The example concludes with the shared control safety argument pattern being instantiated for ALKS. Not shown in this example, but the expectation is that the loss scenario catalogue (i.e. Tables 24, and 25), identified by the analysis, would provide a rich source of potential test scenarios. Thus, enriching the verification strategy beyond that of solely requirements based testing.

E

OXBOTICA WORKSHOP

An STPA based workshop took place with the Oxbotica team on Monday 1st November 2021. The objective of which was twofold: to provide STPA based training to the engineering team, and to evaluate the EVCM and associated method in the context of shared control.

The focus of the workshop was Selenium, Oxbotica's "full stack autonomy system," which when integrated with a suitable road vehicle platform achieves full (SAE Level 4) automation [120]. To give the workshop activities context, the BP Lingen Refinery example was considered [26], with human oversight of the automation being provided by either an External[1] or Remote Operator[2].

The workshop was attended by eight members of the Selenium engineering design team - with six attending in person and two remotely. The delegates had design responsibility for the following aspects of the Selenium system:

- Head of Safety
- Principal Engineer Functional Safety
- Lead Safety Assurance Engineer
- Team Lead Automation Planning and Control
- Principal Product Engineer (Level 4 Autonomy)
- Principal Validation Engineer
- Senior Systems Safety Engineer

1 The human External Operator would be outside of the autonomous vehicle, but with a line-of-sight vantage point to the vehicle while it is operating.

2 Like the External Operator, the Remote Operator is outside the vehicle. However, as the name suggests, they are in a remote location where they may be responsible for providing oversight to multiple vehicles.

- Product Owner (Level 4 Autonomy)

E.1 WORKSHOP STRUCTURE

The workshop used PowerPoint presentation material to convey the ideas presented. To establish a common baseline among the delegates no prior STPA experience was assumed. Consequently, the workshop needed to describe both the *Shared Control* STPA method (as described in Chapter 9) and *Classic* STPA; together with the cognitive principles (e.g. COCOM and ECOM) underpinning the EVCM. The activity of describing the foundation work, together with the limited workshop time available (approximately 4 hours), did undoubtably impact the depth to which concepts could be explored and challenged.

E.2 WORKSHOP OBSERVATIONS

As discussed above, the workshop walked through the steps of the *Shared Control* STPA method, with the below observations made. Sensitive product related information has been excluded from this text, with a more detailed version of this material having been shared with the Oxbotica team as a record of the workshop outcomes.

E.2.1 *Step 1 Scope of the analysis*

STPA Step 1 defines the scope of the analysis. This step began by introducing Selenium [120] in the context of the BP Lingen refinery [26] as the subject of the analysis. With shared control being the focus of the *Shared Control* STPA method introduced, the relationship between Selenium and the External / Remote Operator was the key property of the Selenium system introduced. However, it is worth reiterating that the *Shared Control* STPA method can be used to explore the whole system – and it is not just applicable to shared control.

The system boundary defined excluded the vehicle, vehicle interfaces and vehicle smart actuators. This was done to constrain the subject of the analysis. There is no reason why STPA should not be applied to these system elements also.

The stakeholders, losses and hazards were defined prior to the workshop. These were focused towards stakeholders associated with system safety. It should be noted that the stakeholder list / losses can be expanded further as required. For example, the scope could include topics such as security and financial. The delegates agreed in principle with the list of hazards presented. The difference between hazard H1 *"Ego vehicle does not maintain a minimum safe distance to other road users (includes pedestrians, cyclists and other vehicles)"* and hazard H2 *"Ego vehicle fails to stop for / avoid objects in path"* was discussed. It was observed that H1 might apply to other road users / agents, whereas H2 could relate to objects without agency. However, given these two hazards relate to near identical losses, there is an argument for simply combining them into one hazard. It was noted that "the hazards are defined at a "higher" level than is typical for functional safety."

E.2.2 *Step 2 Model the control structure*

The structure of the STPA CSD was described. That is a hierarchical control structure of controllers, sensors and actuators. Within the control structure each controller enforces constraints on the behaviour of the system. This it does by issuing control actions and monitoring the impact of those control actions – effectively classical close loop control.

The 'pitfalls' associated with creating CSDs were discussed, specifically falling into the trap of going into too much detail too soon. In the *STPA Handbook* Leveson and Thomas stresses the importance of using abstraction to manage complexity. However, the complexity of CSDs typically found in contemporary literature would suggest that analysts struggle to apply abstraction effectively in practice.

The EVCM was introduced as a conceptual model that could be used, in conjunction with the behavioural competency taxonomy, to describe a highly automated vehicle system. A CSD version of the EVCM was presented for use as part of the STPA method.

The behavioural competency taxonomy was reviewed in the context of the BP Lingen refinery. The delegates discussed the behavioural competencies applicable to the refinery ODD. Of those competencies considered applicable to the ODD the delegates then identified the actor responsible in each case. That is, the actor responsible for doing and monitoring each competency, as well as the actor responsible for maintaining vehicle safety. Due to its commercial sensitivity, the delegates' responses to the behavioural competency actor responsibility identification exercise cannot be included with this document. However, the generalised observation made from the delegate responses received are discussed below.

The delegates' individual responses to the behavioural competency actor responsibility identification task were very different. With the benefit of hindsight and a longer workshop duration, more time should have been dedicated to understanding the exercise premise. This would have ensured a common baseline understanding among the delegates from which to measure the 'effectiveness' of the method. As a consequence, it is impossible to know whether the differences observed are as a result of the actor responsibility identification method not being repeatable, or just each delegate having a different mental model of the exercise presented. Certainly with the luxury of time, reflecting on the differences identified would have aligned the team towards a common understanding of the exercise, and may have provided insight into how difficult the delegates actually found the task of ranking agent responsibilities. However, differences in ranking aside, the "behavioural competency table provided a good definition of system capability"; both in terms of the competencies implemented by the system, as well as explicitly capturing those that are not.

In relation to the BP Lingen refinery example, one delegate raised a question about the External Operator - specifically, "where is the External Operator in

the CSD?". The delegates discussed this point and identified that the External Operator was in fact represented throughout the CSD. With "shared control existing anywhere where the responsibility for a behaviour", in the behavioural competency table, "is designated as being 'shared' or the responsibility of the 'human'". Identifying the agent responsibility for each behavioural competency in this way "helped to make the areas of shared control within the system explicit". And by considering human control throughout the EVCM, and hence the CSD, "keeps the human inside the control loop and inside the system." Thus allowing shared control to be "considered more completely."

The delegates considered that the behavioural competency taxonomy was effective at describing the DDT. However, a system like Selenium would also need to implement behavioural competencies not directly related to driving. For example, transferring control between the External Operator and the automation. It was suggested that to describe all behavioural competencies needed to maintain safety, the taxonomy should include competencies such as *manage the transfer of control*. Because "even for Level 4 autonomy" there is a need to manage the transfer of control between the human and the automation.

The observation was made that once complete, the "behavioural competency table provided a good definition of system capability"; both in terms of the competencies implemented by the system, as well as explicitly capturing those that are not. As such, "the behavioural competency table would make a useful addition to the Item Definition." Although on this occasion delegates did not score the agent responsibilities for Selenium the same, which could be seen as a negative for the process presented. The observation was made that the process of alignment, discovering and understanding differences is a valuable attribute of the process.

There was some discussion between delegates about the capability required by Selenium for the BP Lingen refinery ODD, which would have affected the agent responsibility scoring. For example, some delegates identified *reverse the vehicle* as being applicable in the ODD, while others considered this behavioural competency to be out-of-scope for the refinery. Included in the prerequisite

workshop material were definitions for each behavioural competency. However, the delegates were not expected to read the material and commit it to memory beforehand, nor where they explicitly instructed to bring the prerequisite material with them to the workshop. With some or all of the delegates working from memory and the BP Lingen refinery being one of multiple Selenium applications, differences are inevitable.

In hindsight, the notion of agent responsibility types is too complex to expect delegates to assimilate quickly in a workshop setting. To help 'steer' the delegates thinking towards solely *role responsibility*, H. L. A. Hart's seminal *taxonomy of responsibility* was introduced [29, 184]. With the example being given that *role responsibility* involves actor 'S' having a specific duty to bring about 'X' as part of an assigned task, which aligns with the Law Commissions notion of *user in charge* [92], which was probably familiar to the delegates. Dependent on the context, Selenium's *user in charge* could be the driver, an external operator, or a remote operator. The variance in the way the delegates scored *role responsibility* for "do', "monitor" and "achieve safety", and the fact that in some cases "monitor" was scored as "human" while "achieve safety" was identified as "shared" would suggest that different mental models existed amongst the group. This again highlights the importance of dedicating sufficient time to explain and test the delegates' understanding of the exercise premise during the workshop.

The behavioural competency interaction diagram (similar to the ALKS behavioural competency interaction diagram shown in Figure 35 on page 282) was also presented as a visual tool that can be used to identify interactions between competencies. The delegates engaged less with the behavioural competency interaction diagram, so its usefulness as a tool to reason about shared control was inconclusive. The reason for this was probably due in part to the limited time available to practice with the concept. But, also the paper template presented was perhaps a little inflexible compared to creating an electronic interaction diagram in a suitable drawing tool. Which is how the author has approached the creation of behavioural competency interaction diagrams in the past.

E.2.3 *Step 3 Identify unsafe control actions*

The definition of an 'unsafe control action' as "a control action that, in a particular context and worst-case environment, will lead to a hazard" was introduced. The guide word led process for identifying unsafe control actions from control actions was described. The group used the control action *target trajectory* to discuss the unsafe control actions process step; which for Selenium is a vector quantity describing a few seconds of lateral and longitudinal vehicle movement. This discussion highlighted that – as with other analysis techniques like HAZOP – a level of interpretation is needed when applying the guide words to the control actions. Thus, for transparency and reuse the importance of documenting the rationale / interpretation of each guide word was highlighted.

E.2.4 *Step 4 Identify loss scenarios*

The *Classic* STPA method describes two situations where an unsafe control action might become a loss scenario. In this context a loss scenario describes a scenario in which causal factors can lead to the unsafe control action and to hazards. The *Shared Control* STPA method described during the workshop includes two further situations, relevant to shared control, where an unsafe control action may include the causal factors necessary to result in a hazard. During the workshop these four situations where described pictorially, and are summarised as:

- Scenarios that lead to unsafe control actions due to unsafe controller behaviours or due to inadequate feedback and information to the controller.
- Scenarios in which control actions are executed improperly or not executed. This might be due to causal factors in the control path or in the controlled process itself.
- Scenarios in which an unsafe control action results due to an external event not being evaluated correctly or because situational awareness is

incorrectly maintained. In such scenarios, causal factors might impact an agents reaction times, understanding of the system behaviour, or of the environment.

- Scenarios in which the agent chooses the wrong control action to perform. This includes considering what might cause an agent to choose the wrong action and the time needed to make such a decision.

As background to scenario types 3 and 4 above, Erik Hollnagel's COCOM and ECOM were introduced. Although intended to represent the "cyclical model of human actions" it's potential applicability to systems that include machine learning algorithms was hypothesised.

Again, the workshop did not provide the 'right' environment in which to allow the Team to practice using the loss scenario questions to identify causes of unsafe control actions. However, the expectation is that when the delegates undertake their own *thought experiments* they will be able to use the loss scenario identification questions in combination with the EVCM to successfully analyse the unsafe control actions identified during STPA Step 3. This is the subject of on-going research with Oxbotica.

Finally the opportunity for the identified loss scenarios to provide a rich knowledge base from which to derive further system safety requirements and test scenarios was discussed.

E.3 SUMMARY

E.3.1 *The EVCM and the hierarchy of control*

The group felt that as a conceptual model the EVCM structure was reflective of a highly automated system, such as Selenium. It was felt that the hierarchy of control levels (i.e. control, manoeuvring and strategy) provided a "good framework in which to describe autonomous systems." An advantage

compared to typical functional safety techniques that tend to be at a "too lower level of detail for AI systems."

Having seen how the EVCM can be used to describe Selenium, Oxbotica plan to continue using the EVCM as the CSD for their ongoing STPA analysis. By aligning Selenium's architecture to the control, manoeuvring and strategy hierarchical levels Oxbotica expects this will aid the safety requirements elicitation process. That is, allowing safety requirements, identified during STPA Steps 3 and 4, to be directly allocated to the appropriate control layer of the system architecture.

E.3.2 *Behavioural competency taxonomy*

The group could see how the behavioural competency taxonomy could be used to describe the system behaviour. Both as a description of the behaviours implemented by the system, but also as an indication of the behaviours a given ODD might demand. In all likelihood the taxonomy is complete for the DDT. However, to describe all behavioural competencies needed to maintain safety, the taxonomy should include competencies such as *manage the transfer of control*. Because "even for Level 4 autonomy there is a need to manage the transfer of control between the human and the automation."

The team observed that once complete, the behavioural competency table provided a good definition of system capability; both in terms of the competencies implemented by the system, as well as explicitly capturing those that are not. As such, it was felt that the behavioural competency table would make a "useful addition to the Item Definition."

Using the behavioural competency taxonomy to identify behaviours that are either undertaken by the human operator, or that are shared between the human operator and the automation, made the areas of shared control within the system explicit. It was felt that this approach dealt with shared control more holistically and "helped to keep the human operator more inside the system and a part of the control system."

E.3.3 *Next steps*

The team intends to use the techniques learned during the workshop to explore Selenium hazard causes, both in the context of BP Lingen and the other partner applications currently under development. Regular meetings are scheduled between Oxbotica and the author. These are intended to provide the Oxbotica team with an opportunity to seek further support regarding the method, but also for the author and the University of York to gain further feedback on the method's effectiveness.

F

EVALUATION STRATEGY

The evaluation strategy first described in Chapter 11 is represented within this appendix pictorially, using GSN. For the accompanying narrative, the reader is directed to Chapter 11, page 163.

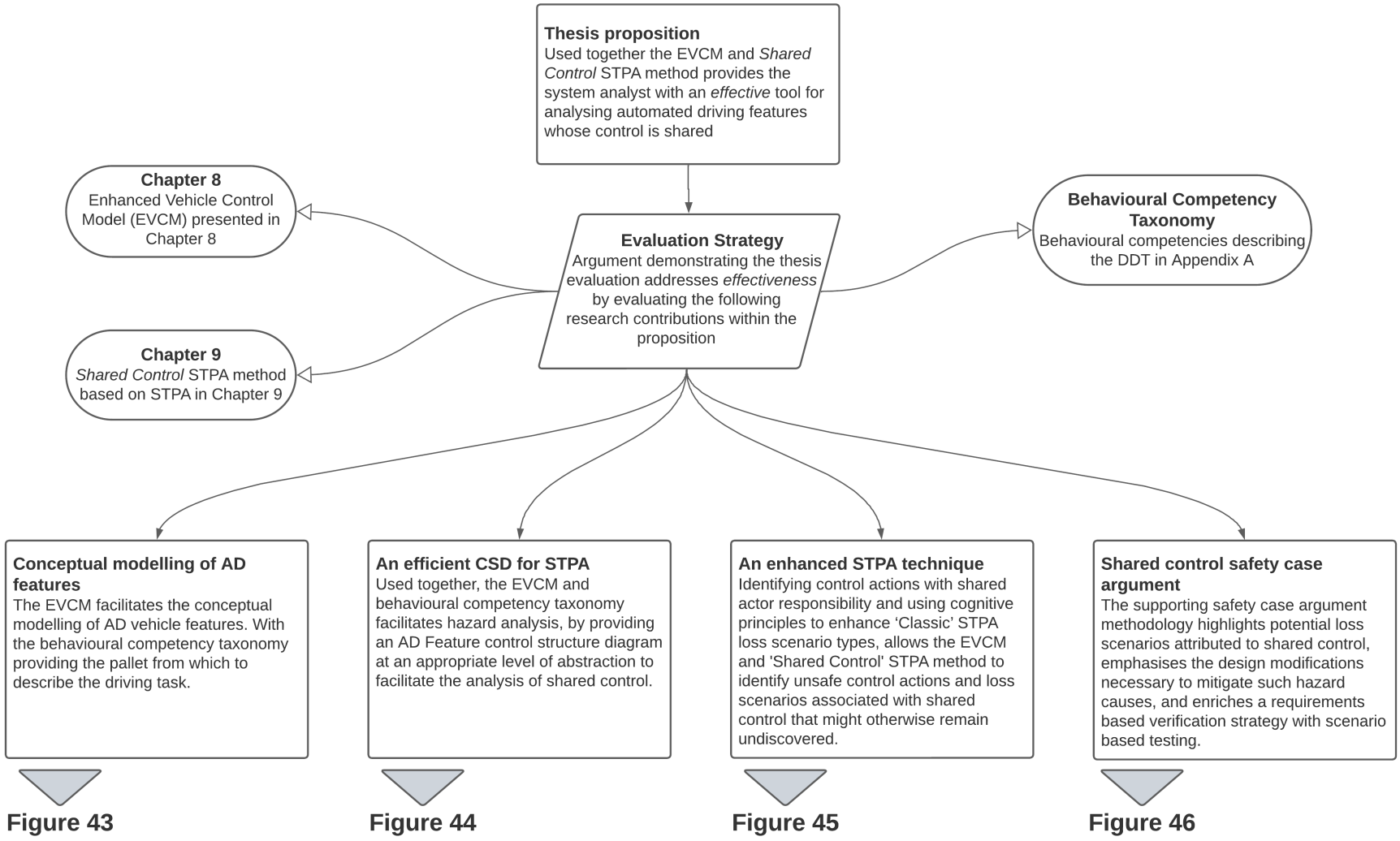


Figure 42: Thesis proposition evaluation

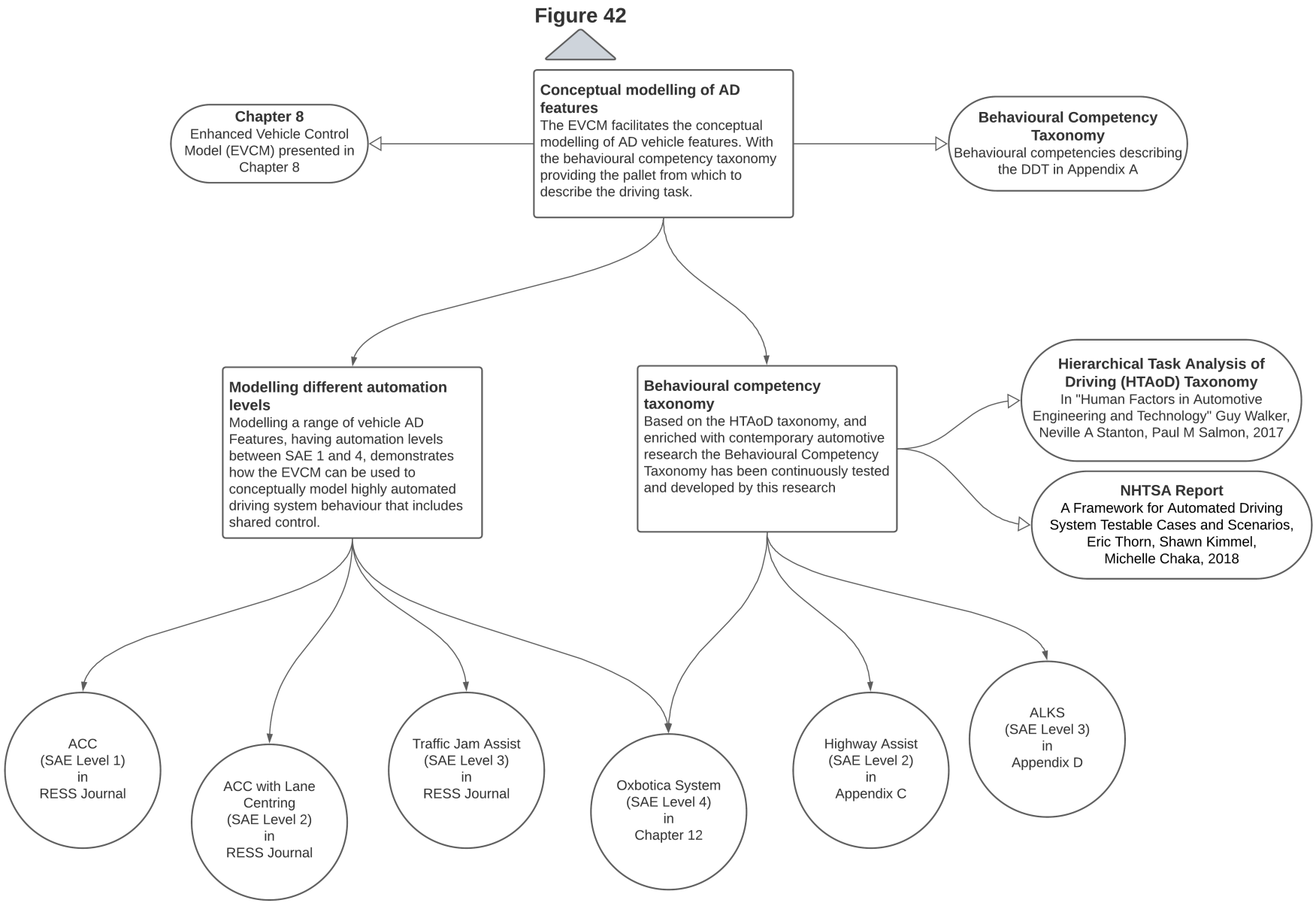


Figure 43: AD feature conceptually modelling evaluation

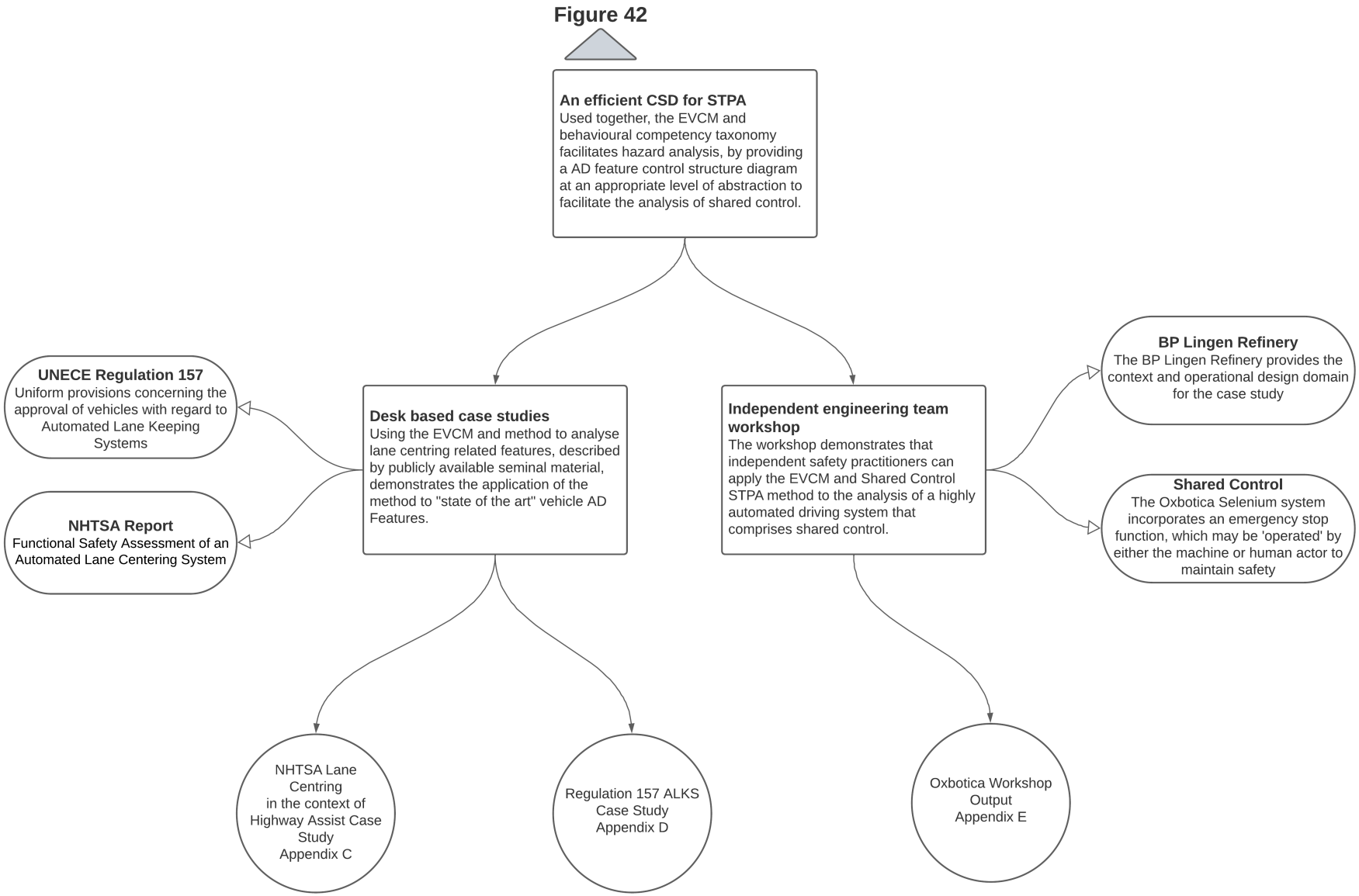


Figure 44: Evaluating the EVCM as a CSD for STPA

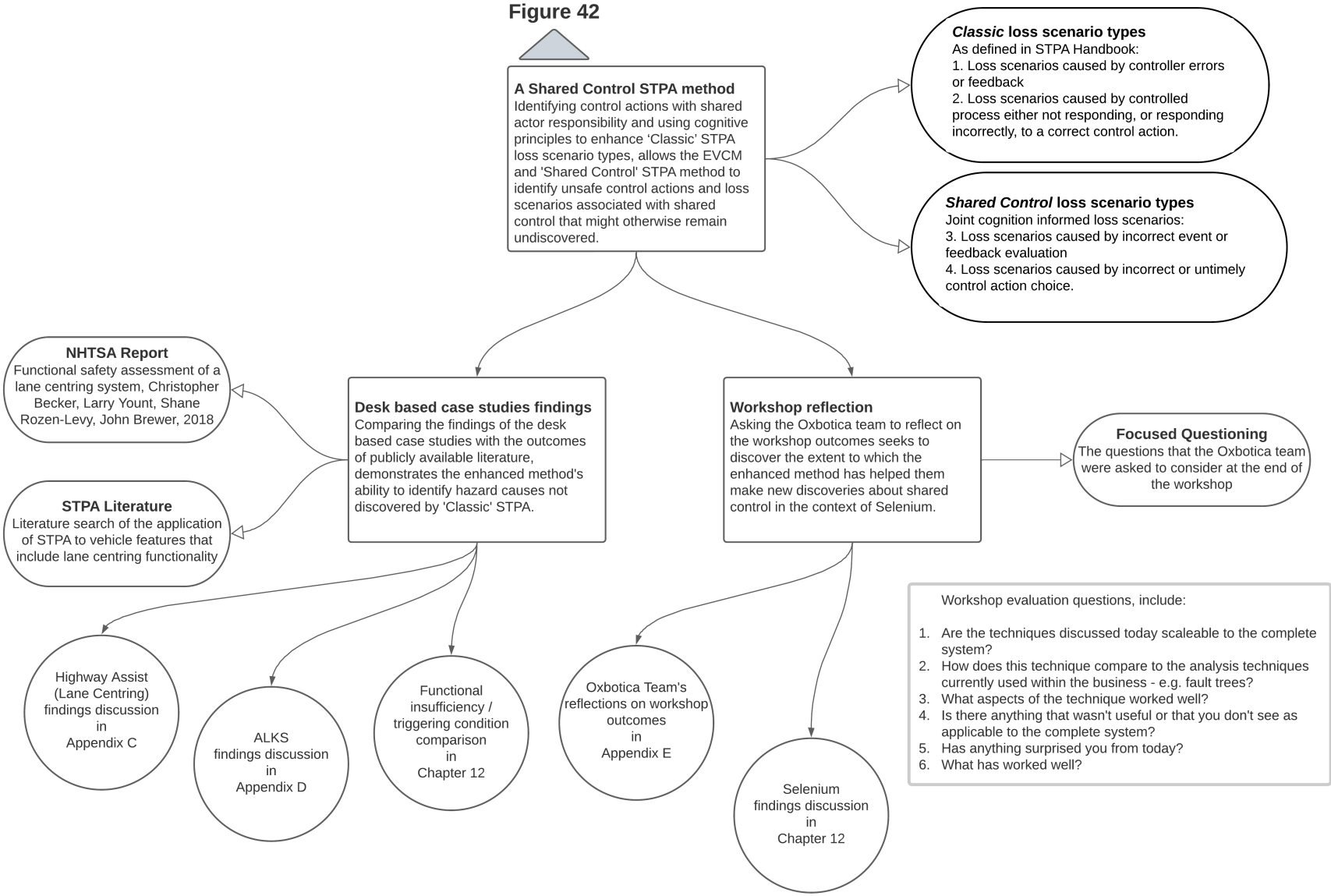


Figure 45: Shared control STPA method evaluation

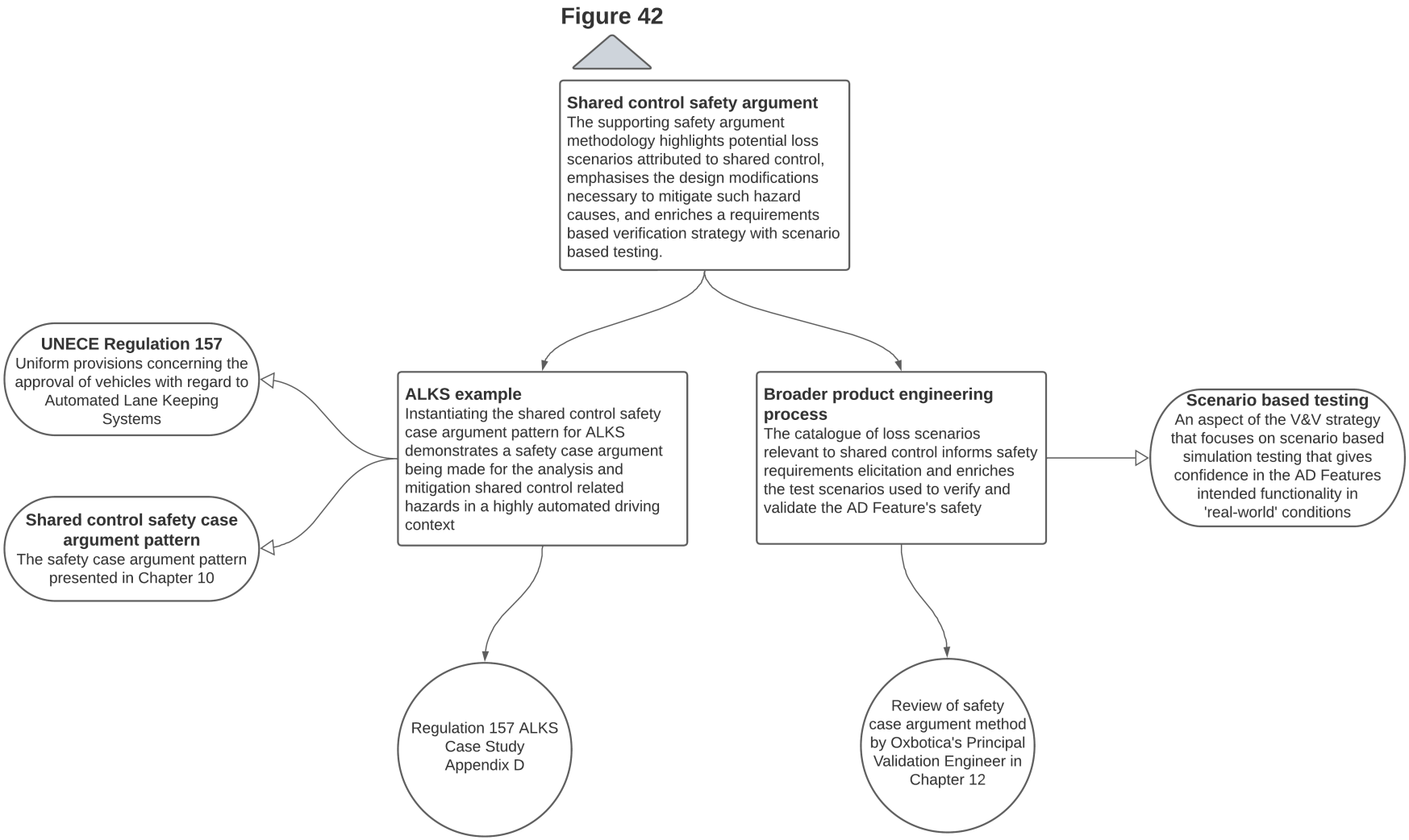


Figure 46: Shared control safety case argument evaluation

GLOSSARY

ABS Anti-lock Braking System

ACC Active Cruise Control

AD Automated Driving

ADAS Advanced Driver Assistance System

AEB Automatic Emergency Braking

ALARP As Low As Reasonably Practicable

ALC Automated Lane Centring

ALKS Automatic Lane Keeping System

AMLAS Assurance of Machine Learning for use in Autonomous Systems

ASIL Automotive Safety Integrity Level

BSM Blind Spot Monitoring

CAST Controllability of Automotive Safety Targets

CAST Causal Analysis based on Systems Theory

CBA Cost Benefit Analysis

COCOM Contextual Control Model

CSD Control Structure Diagram

CSE Cognitive Systems Engineering

DDT Dynamic Driving Task

DRIVE Dedicated Road Infrastructure for Vehicle Safety in Europe

DSA Distributed Situation Awareness

ECOM Extended Control Model

E/E Electrical/Electronic

E/E/PE Electrical/Electronic/Programming Electronic

EAST Event Analysis of Systemic Team-work

ESC Electronic Stability Control

EVCM Enhanced Vehicle Control Model

EUC Equipment Under Consideration

Euro NCAP European New Car Assessment Programme

FFA Functional Failure Analysis

FMEA Failure Mode Effects Analysis

FRAM Functional Resonance Analysis Method

FTA Fault Tree Analysis

GPS Global Positioning System

GSN Goal Structuring Notation

HARA Hazard Analysis and Risk Assessment

HCM Hierarchical Control Model

HAS Highway Assist System

HAZOP Hazard and Operability Study

HCM Hierarchical Control Model

HMC Human Machine Collaboration

HMI Human Machine Interface

HMI Human Machine Interaction

HMT Human Machine Teaming

HTAoD Hierarchical Task Analysis of Driving

IEC International Electrotechnical Commission

ISO International Organization for Standardization

JCS Joint Cognitive Systems

LiDAR Light Detection And Ranging

LKA Lane Keeping Assistance

LOPA Layers of Protection Analysis

MISRA Motor Industry Software Reliability Association

MMR Measles, Mumps, and Rubella

MRM Minimum Risk Manoeuvre

NAT Normal Accident Theory

NHS National Health Service

NHTSA National Highway Traffic Safety Administration

ODD Operational Design Domain

OEDR Object and Event Detection and Response

OEM Original Equipment Manufacturer

OEMs Original Equipment Manufacturers

OODA Observe Orient Decide Act

PAS Publicly Available Specification

QD Qualifying Dissertation

QM Quality Management

RADAR RAdio Detection And Ranging

RFID Radio Frequency Identification

SA Situation Awareness

SAE Society of Automotive Engineers

SIL Safety Integrity Level

SOTIF Safety Of The Intended Functionality

SPA Sense Plan Act

SRS Secondary Restraints System

STAMP Systems Theoretic Accident Model and Processes

STOE-KRS Strategic Tactical Operational Executional – Knowledge Rule Skill

STPA Systems-Theoretic Process Analysis

SUDA Sense Understand Decide Act

TJA Traffic Jam Assist

UCA Unsafe Control Action

UNECE United Nations Economic Commission for Europe

VCM Vehicle Control Model

WHO World Health Organisation

BIBLIOGRAPHY

[1] David A. Abbink, Tom Carlson, Mark Mulder, Joost C. F. de Winter, Farzad Aminravan, Tricia L. Gibo and Erwin R. Boer. "A Topology of Shared Control Systems—Finding Common Ground in Diversity". In: *IEEE Transactions on Human-Machine Systems* 48.5 (2018), pp. 509–525. ISSN: 2168-2291 2168-2305. DOI: 10.1109/thms.2018.2791570.

[2] Asim Abdulkhaleq, Markus Baumeister, Hagen Böhmert and Stefan Wagner. "Missing no Interaction—Using STPA for Identifying Hazardous Interactions of Automated Driving Systems". In: *International Journal of Safety Science* 02.01 (2018), pp. 115–124. ISSN: 23716312 23716320. DOI: 10.24900/ijss/0201115124.2018.0301.

[3] Masato Abe. *Vehicle handling dynamics: theory and application*. Butterworth-Heinemann, 2015. ISBN: 0081003730.

[4] Sara Abidi Nasri. "Application of the STPA methodology to an automotive system in compliance with ISO26262". Thesis. 2018.

[5] National Highway Traffic Safety Administration. "Preliminary statement of policy concerning automated vehicles". In: *Washington, DC* (2013).

[6] Alfa Romeo USA. *Highway Assist System/Traffic Jam Assist – How To – 2020 Alfa Romeo Stelvio*. 2020. URL: https://www.youtube.com/watch?v=ywj7GCtF9XM (visited on 29th December 2020).

[7] C. E. Althaus. "A disciplinary perspective on the epistemological status of risk". In: *Risk Anal* 25.3 (2005), pp. 567–88. ISSN: 0272-4332 (Print) 0272-4332 (Linking). DOI: 10.1111/j.1539-6924.2005.00625.x. URL: http://www.ncbi.nlm.nih.gov/pubmed/16022691.

[8] J Joseph Antony and M Suchetha. "Vision Based Vehicle Detection: A Literature Review". In: *International Journal of Applied Engineering Research* 11.5 (2016), pp. 3128–3133. ISSN: 0973-4562.

[9] A. Arlow, C. Duffy and J. McDermid. "The Safety Programme For The Specification Of The Active Traffic Management Control System". In: 2006.

[10] Teresa Ashe. *How the media report scientific risk and uncertainty: a review of the literature*. Report. Reuters Institute for the Study of Journalism, 2013.

[11] Assuring Autonomy International Programme. *Assurance of Machine Learning for use in Autonomous Systems*. 2021. URL: https://www.york.ac.uk/assuring-autonomy/guidance/amlas/.

[12] Automobile Association. *The evolution of car safety features: From windscreen wipers to crash tests and pedestrian protection*. 2019. URL: https://www.theaa.com/breakdown-cover/advice/evolution-of-car-safety-features (visited on 27th May 2019).

[13] Terje Aven. "On how to define, understand and describe risk". In: *Reliability Engineering & System Safety* 95.6 (2010), pp. 623–631. ISSN: 09518320. DOI: 10.1016/j.ress.2010.01.011.

[14] Terje Aven. "Risk assessment and risk management: Review of recent advances on their foundation". In: *European Journal of Operational Research* (2015). ISSN: 03772217. DOI: 10.1016/j.ejor.2015.12.023.

[15] Terje Aven and Ortwin Renn. "On risk defined as an event where the outcome is uncertain". In: *Journal of Risk Research* 12.1 (2009), pp. 1–11. ISSN: 1366-9877 1466-4461. DOI: 10.1080/13669870802488883.

[16] S. Ben Ayed, H. Trichili and A. M. Alimi. "Data fusion architectures: A survey and comparison". In: *2015 15th International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 277–282. DOI: 10.1109/ISDA.2015.7489238.

[17] Lisanne Bainbridge. "Ironies of automation". In: *Automatica* 19.6 (1983), pp. 775–779. ISSN: 0005-1098. DOI: http://dx.doi.org/10.1016/0005-

1098(83)90046-8. URL: http://www.sciencedirect.com/science/article/pii/0005109883900468.

[18] Victoria A. Banks, Alexander Eriksson, Jim O'Donoghue and Neville A. Stanton. "Is partially automated driving a bad idea? Observations from an on-road study". In: *Applied Ergonomics* 68 (2018), pp. 138–145. ISSN: 0003-6870. DOI: https://doi.org/10.1016/j.apergo.2017.11.010. URL: http://www.sciencedirect.com/science/article/pii/S0003687017302594.

[19] Victoria A Banks and Neville A Stanton. "Analysis of driver roles: Modelling the changing role of the driver in automated driving systems using EAST". In: *Theoretical Issues in Ergonomics Science* (2017).

[20] Christopher Becker, Larry Yount, Shane Rozen-Levy and John Brewer. *Functional safety assessment of a lane centring system - appendices*. Journal Article. National Highway Traffic Safety Administration, 2018.

[21] Christopher Becker, Larry Yount, Shane Rozen-Levy and John Brewer. *Functional Safety Assessment of an Automated Lane Centering System*. Journal Article. National Highway Traffic Safety Administration, 2018.

[22] Thomas Berger and Timo Reis. "Controllability of linear differential–algebraic systems—a survey". In: *Surveys in Differential–Algebraic Equations I*. Springer, 2013, pp. 1–61. ISBN: 3642349277.

[23] John Birch, David Blackburn, John Botham, Ibrahim Habli, David Higham, **Helen Monkhouse**, Gareth Price, Norina Ratiu and Roger Rivett. "A Structured Argument for Assuring Safety of the Intended Functionality (SOTIF)". In: *Computer Safety, Reliability, and Security. SAFECOMP Workshops*. Lecture Notes in Computer Science. 2020. Chap. Chapter 31, pp. 408–414. ISBN: 978-3-030-55582-5. DOI: 10.1007/978-3-030-55583-2_31.

[24] John Birch, Roger Rivett, Ibrahim Habli, Ben Bradshaw, John Botham, Dave Higham, Peter Jesty, **Helen Monkhouse** and Robert Palin. "Safety cases and their role in ISO 26262 functional safety assessment". In: *Computer Safety, Reliability, and Security*. Springer, 2013, pp. 154–165. ISBN: 3642407927.

[25] Born to Drive. *Alfa Romeo Giulia and Stelvio Highway Assist Level 2 Autonomous Driving: How It Works*. 2020. URL: https://www.youtube.com/watch?v=wrJ8flUAvnI (visited on 29th December 2020).

[26] BP PLC. *BP and Oxbotica complete industry-first autonomous vehicle trial at German refinery*. URL: https://www.bp.com/en/global/corporate/news-and-insights/press-releases/bp-and-oxbotica-complete-industry-first-autonomous-vehicle-trial-at-german-refinery.html (visited on 2nd August 2021).

[27] Simon Burton. *Blog post: Robust and resilient. Designing safe automated driving systems*. Tech. rep. AAIP University of York, June 2020. URL: https://www.york.ac.uk/assuring-autonomy/news/blog/safety-highly-automated-driving-robust-resilient/ (visited on 9th January 2022).

[28] P. Carlo Cacciabue. "Modelling Driver Behaviour In Automotive Environments". In: (2007). ISSN: 10: 1-84628-617-4.

[29] Peter Cane. "Role Responsibility". In: *The Journal of Ethics* 20.1-3 (2016), pp. 279–298. ISSN: 1382-4554 1572-8609. DOI: 10.1007/s10892-016-9235-8.

[30] O. Carsten, F. C. H. Lai, Y. Barnard, A. H. Jamson and N. Merat. "Control Task Substitution in Semiautomated Driving: Does It Matter What Aspects Are Automated?" In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 54.5 (2012), pp. 747–761. ISSN: 0018-7208 1547-8181. DOI: 10.1177/0018720812460246.

[31] Diogo Silva Castilho, Ligia M. S. Urbina and Donizeti de Andrade. "STPA for continuous controls: A flight testing study of aircraft cross-wind takeoffs". In: *Safety Science* 108 (2018), pp. 129–139. ISSN: 09257535. DOI: 10.1016/j.ssci.2018.04.013.

[32] CEDR's Secretariat General. *Forgiving roadside design guide*. Tech. rep. Conference of European Directors of Roads, November 2012. URL: https://www.cedr.eu/download/Publications/2013/T10_Forgiving_roadsides.pdf (visited on 2nd January 2022).

[33] Georgios Chrysakis, Helen Monkhouse and Stratis Kanarachos. "Vehicle controllability assessment using detailed multibody vehicle simulations". In: *FAST-zero'15: 3rd International Symposium on Future Active Safety Technology Toward zero traffic accidents, 2015*.

[34] George E Cooper and Robert P Harper Jr. *The use of pilot rating in the evaluation of aircraft handling qualities*. Report. DTIC Document, 1969.

[35] Vincent T. Covello. "The perception of technological risks: A literature review". In: *Technological Forecasting and Social Change* 23.4 (1983), pp. 285–297. ISSN: 0040-1625. DOI: http://dx.doi.org/10.1016/0040-1625(83)90032-X. URL: http://www.sciencedirect.com/science/article/pii/004016258390032X.

[36] M. Da Lio, A. Mazzalai, K. Gurney and A. Saroldi. "Biologically Guided Driver Modeling: the Stop Behavior of Human Car Drivers". In: *IEEE Transactions on Intelligent Transportation Systems* 19.8 (2018), pp. 2454–2469. ISSN: 1524-9050. DOI: 10.1109/tits.2017.2751526. URL: %3CGo%20to%20ISI%3E://WOS:000440852900007.

[37] K. De Groot and R. Thurik. "Disentangling Risk and Uncertainty: When Risk-Taking Measures Are Not About Risk". In: *Front Psychol* 9 (2018), p. 2194. ISSN: 1664-1078 (Print) 1664-1078 (Linking). DOI: 10.3389/fpsyg.2018.02194. URL: https://www.ncbi.nlm.nih.gov/pubmed/30498464.

[38] Andrew Del-Colle. *This Cadillac Concept Had Radar Crash-Avoidance in 1959*. October 2015. URL: https://www.roadandtrack.com/car-culture/news/a27132/the-first-radar-based-crash-avoidance-system-was-in-a-1959-concept-car/ (visited on 21st January 2022).

[39] Department of Transport, UK Government. *The Highway Code*. URL: https://www.gov.uk/guidance/the-highway-code (visited on 3rd April 2022).

[40] J. Dillmann, R. J. R. den Hartigh, C. M. Kurpiers, F. K. Raisch, D. de Waard and R. F. A. Cox. "Keeping the driver in the loop in conditionally automated driving: A perception-action theory approach". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 79 (2021), pp. 49–62. ISSN: 13698478. DOI: 10.1016/j.trf.2021.03.003.

[41] *DO-178C Software Considerations in Airborne Systems and Equipment Certification*. RTCA & EUROCAE, 2012.

[42] Lakshmi S Dutt and Mathew Kurian. "Handling of uncertainty-A Survey". In: *International Journal of Scientific and Research Publications* 3.1 (2013), pp. 1–4.

[43] EGAS Workgroup. *Standardized E-Gas Monitoring Concept for Gasoline and Diesel Engine Control Units*. EGAS Workgroup, May 2013.

[44] M Ellims, H Monkhouse and A Lyon. "ISO 26262: Experience applying part 3 to an in-wheel electric motor". In: *6th IET International Conference on Systems Safety*. The Institution of Engineering and Technology, 2011, pp. 1–8.

[45] Michael Ellims, Helen Elizabeth Monkhouse, Damian Harty and Teena Gade. "Using Vehicle Simulation to Investigate Controllability". In: *SAE International Journal of Alternative Powertrains* 2.2013-01-0180 (2013), pp. 18–36. ISSN: 2167-4205.

[46] Rune Elvik, Truls Vaa, Alena Hoye and Michael Sorensen. *The handbook of road safety measures*. Emerald Group Publishing, 2009. URL: https://www.emerald.com/insight/publication/doi/10.1108/9781848552517 (visited on 28th July 2022).

[47] Mica R. Endsley. "Toward a Theory of Situation Awareness in Dynamic Systems". In: *Human Factors* 37.1 (1995), pp. 32–64. DOI: 10.1518/001872095779049543. URL: https://journals.sagepub.com/doi/abs/10.1518/001872095779049543.

[48] Johan Engström, Jonas Bärgman, Daniel Nilsson, Bobbie Seppelt, Gustav Markkula, Giulio Bianchi Piccinini and Trent Victor. "Great expectations: a predictive processing account of automobile driving". In: *Theoretical Issues in Ergonomics Science* 19.2 (2017), pp. 156–194. ISSN: 1463-922X 1464-536X. DOI: 10.1080/1463922x.2017.1306148.

[49] Johan Engström and Erik Hollnagel. "A General Conceptual Framework for Modelling Behavioural Effects of Driver Support Functions". In: *Modelling Driver Behaviour in Automotive Environments: Critical Issues in*

Driver Interactions with Intelligent Transport Systems. London: Springer London, 2007. Chap. 4, pp. 61–84. ISBN: 978-1-84628-618-6.

[50] A. Eriksson and N. A. Stanton. "Takeover Time in Highly Automated Vehicles: Noncritical Transitions to and From Manual Control". In: *Hum Factors* 59.4 (2017), pp. 689–705. ISSN: 1547-8181 (Electronic) 0018-7208 (Linking). DOI: 10.1177/0018720816685832. URL: https://www.ncbi.nlm.nih.gov/pubmed/28124573.

[51] Alexander Eriksson, Sebastiaan M. Petermeijer, Markus Zimmermann, Joost C. F. de Winter, Klaus J. Bengler and Neville A. Stanton. "Rolling Out the Red (and Green) Carpet: Supporting Driver Decision Making in Automation-to-Manual Transitions". In: *IEEE Transactions on Human-Machine Systems* 49.1 (2019), pp. 20–31. ISSN: 2168-2291 2168-2305. DOI: 10.1109/thms.2018.2883862.

[52] European New Car Assessment Programme. *How to read the stars*. URL: https://www.euroncap.com/en/about-euro-ncap/ (visited on 7th January 2022).

[53] European New Car Assessment Programme. *Road Map 2025 – In Pursuit of Vision Zero*. Tech. rep. September 2017. URL: https://cdn.euroncap.com/media/30700/euroncap-roadmap-2025-v4.pdf (visited on 8th January 2022).

[54] R Factor. "The influence of social characteristics on drivers' involvement in traffic accidents". PhD thesis. Faulty of Law, The Hebrew University of Jerusalem, 2008.

[55] Baruch Fischhoff, Paul Slovic, Sarah Lichtenstein, Stephen Read and Barbara Combs. "How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits". In: *Policy Sciences* 9.2 (1978), pp. 127–152. ISSN: 1573-0891. DOI: 10.1007/bf00143739.

[56] Frank Flemisch, Johann Kelsch, Christian Löper, Anna Schieben and Julian Schindler. "Automation spectrum, inner/outer compatibility and other potentially useful human factors concepts for assistance and automation". In: *Human Factors for assistance and automation (2008)* (2008), pp. 1–16.

[57] Megan Elizabeth France. "Engineering for Humans: A New Extension to STPA". MA thesis. Massachusetts Institute of Technology, June 2017.

[58] Functional Safety Committee. *J2980 — Considerations for ISO 26262 Hazard Analysis Classification*. SAE International, April 2018. DOI: https://doi.org/10.4271/J2980_201804.

[59] Tom Gasser et al. *Legal Consequences of an Increase in Vehicle Automation*. Report. Report on the research project F1100.5409013.01of the operational programme of the Federal Highway Research Institute, 2014.

[60] Geneva: World Health Organization. *Save LIVES – A road safety technical package*. Tech. rep. 2017. URL: http://apps.who.int/iris/bitstream/handle/10665/255199/9789241511704-eng.pdf?sequence=1&isAllowed=y (visited on 8th January 2022).

[61] Thomas D. Gillespie. *Fundamentals of Vehicle Dynamics*. SAE International, 1992.

[62] Cody Godwin. *Tesla's Autopilot 'tricked' to operate without driver*. April 2021. URL: https://www.bbc.co.uk/news/technology-56854417 (visited on 21st January 2022).

[63] Eryn Grant, Paul M. Salmon, Nicholas J. Stevens, Natassia Goode and Gemma J. Read. "Back to the future: What do accident causation models tell us about accident prediction?" In: *Safety Science* 104 (2018), pp. 99–109. ISSN: 09257535. DOI: 10.1016/j.ssci.2017.12.018.

[64] Anton Green, Dianne Gardner and Stephen Legg. "An exploration of the emotional experience of BASE jumping". In: *Sport in Society* (2020), pp. 1–15. DOI: 10.1080/17430437.2020.1807954. eprint: https://doi.org/10.1080/17430437.2020.1807954. URL: https://doi.org/10.1080/17430437.2020.1807954.

[65] Ibrahim Habli, Richard Hawkins and Tim Kelly. "Software safety: relating software assurance and software integrity". In: *International Journal of Critical Computer-Based Systems* 1.4 (2010), pp. 364–383. ISSN: 1757-8779.

[66] Ibrahim Habli, Ireri Ibarra, Roger S Rivett and Tim Kelly. *Model-based assurance for justifying automotive functional safety*. Report 0148-7191. SAE Technical Paper, 2010.

[67] William Haddon, Edward A. Suchman and David Klein. *Accident Research: Methods and Approaches*. New York: Harper & Row, 1964.

[68] David L Hall and James Llinas. "An introduction to multisensor data fusion". In: *Proceedings of the IEEE* 85.1 (1997), pp. 6–23. ISSN: 0018-9219.

[69] P. A. Hancock et al. "Challenges to Human Drivers in Increasingly Automated Vehicles". In: *Hum Factors* 62.2 (2020), pp. 310–328. ISSN: 1547-8181 (Electronic) 0018-7208 (Linking). DOI: 10.1177/0018720819900402. URL: https://www.ncbi.nlm.nih.gov/pubmed/32022583.

[70] S. O. Hansson and T. Aven. "Is risk analysis scientific?" In: *Risk Anal* 34.7 (2014), pp. 1173–83. ISSN: 1539-6924 (Electronic) 0272-4332 (Linking). DOI: 10.1111/risa.12230. URL: https://www.ncbi.nlm.nih.gov/pubmed/24919396.

[71] Herbert Lionel Adolphus Hart. *Punishment and responsibility: Essays in the philosophy of law*. Oxford University Press, 2008.

[72] Health and Safety Excecutive. *ALARP At a Glance*. 2017. URL: http://www.hse.gov.uk/risk/theory/alarpglance.htm (visited on 22nd January 2017).

[73] C. Herdtweck and C. Curio. "Experts of probabilistic flow subspaces for robust monocular odometry in urban areas". In: *2012 IEEE Intelligent Vehicles Symposium*, pp. 661–667. DOI: {10.1109/IVS.2012.6232238}.

[74] Debra S. Herrmann. *Software Safety and Reliability: techniques, approaches, and standards of key industrial sectors*. IEEE Computer Society, 1999.

[75] *History of road safety, The Highway Code and the driving test*. February 2019. URL: https://www.gov.uk/government/publications/history-of-road-safety-and-the-driving-test/history-of-road-safety-the-highway-code-and-the-driving-test (visited on 1st January 2022).

[76] Erik Hollnagel. *Cognitive Systems Engineering: RIP*. URL: https://erikhollnagel.com/ideas/cognitive-systems-engineering.html (visited on 29th October 2021).

[77] Erik Hollnagel. *Contextual Control Model (COCOM)*. URL: https://erikhollnagel.com/onewebmedia/COCOM.pdf (visited on 29th October 2017).

[78] Erik Hollnagel. *Extended Control Model (ECOM)*. URL: https://erikhollnagel.com/onewebmedia/ECOM.pdf (visited on 29th October 2021).

[79] Erik Hollnagel. "FRAM-1 Understanding Accidents". In: *University of Southern Denmark Presentation* (2011). URL: http://www.functionalresonance.com/FRAM-1_understanding_accidents.pdf (visited on 10th February 2022).

[80] Erik Hollnagel. *FRAM: the Functional Resonance Analysis Method*. Ashgate Publishing Ltd, 2012.

[81] Erik Hollnagel and David D. Wood. *Joint Cognitive Systems – Foundations of Cognitive Systems Engineering*. CRC Press, 2005.

[82] *Institute for Traffic Accident Research and Data Analysis*. 2022. URL: https://www.itarda.or.jp/english (visited on 23rd January 2022).

[83] International Electrotechnical Commission and others. "Functional safety of electrical/electronic/programmable electronic safety related systems". In: *IEC 61508* Edition 2 (2010).

[84] Ward Nicholas J. "Automation of task processes: An example of intelligent transportation systems". In: *Human Factors and Ergonomics in Manufacturing & Service Industries* 10.4 (2000), pp. 395–408. DOI: doi:10.1002/1520-6564(200023)10.

[85] P. H. Jesty, T. F. Buckley and M. M. West. "The development of safe advanced road transport telematic software". In: *Microprocessors and Microsystems* 17.1 (1993), pp. 37–46. ISSN: 0141-9331.

[86] Peter H Jesty, Keith M Hobley, Richard Evans and Ian Kendall. "Safety analysis of vehicle-based systems". In: *Proceedings of the 8th Safety-critical Systems Symposium*. Citeseer, pp. 90–110.

[87] X. Jin, G. Yin and N. Chen. "Advanced Estimation Techniques for Vehicle System Dynamic State: A Survey". In: *Sensors (Basel)* 19.19 (2019). ISSN: 1424-8220 (Electronic) 1424-8220 (Linking). DOI: 10.3390/s19194289. URL: https://www.ncbi.nlm.nih.gov/pubmed/31623345.

[88] Somya Joshi, Thierry Bellet, Vanessa Bodard and Angelos Amditis. "Perceptions of risk and control: Understanding acceptance of advanced driver assistance systems". In: *Human-Computer Interaction–INTERACT 2009*. Springer, 2009, pp. 524–527. ISBN: 3642036546.

[89] J. Klamka. "Controllability of dynamical systems. A survey". In: *Bulletin of the Polish Academy of Science* 61.2 (2013), p. 335. DOI: 10.2478/bpasts-2013-0031.

[90] G. K. Kountouriotis and N. Merat. "Leading to distraction: Driver distraction, lead car, and road environment". In: *Accid Anal Prev* 89 (2016), pp. 22–30. ISSN: 1879-2057 (Electronic) 0001-4575 (Linking). DOI: 10.1016/j.aap.2015.12.027. URL: http://www.ncbi.nlm.nih.gov/pubmed/26785327.

[91] Danica Kragic, Joakim Gustafson, Hakan Karaoguz, Patric Jensfelt and Robert Krug. "Interactive, Collaborative Robots: Challenges and Opportunities". In: *IJCAI*, pp. 18–25.

[92] Law Commission. *Automated Vehicles: A joint preliminary consultation paper*. Tech. rep. Consultation Paper No 240. Law Commission of England and Wales, 2018.

[93] John D. Lee. "Dynamics of Driver Distraction: The process of engaging and disengaging". In: *Annals of Advances in Automotive Medicine* 58 (2014), pp. 24–32. ISSN: 1943-2461. URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001670/.

[94] Nancey G Leveson. *Safeware: System Safety and Computers*. Addison-Wesley, 1995.

[95] Nancy Leveson. "A new accident model for engineering safer systems". In: *Safety Science* 42.4 (2004), pp. 237–270. ISSN: 09257535. DOI: 10.1016/s0925-7535(03)00047-x.

[96] Nancy G Leveson and John P Thomas. *STPA Handbook*. Nancy Leveson, March 2018.

[97] Anders Lindgren and Fang Chen. "State of the art analysis: An overview of advanced driver assistance systems (ADAS) and possible human factors issues". In: *Human factors and economics aspects on safety* 38 (2006), p. 50.

[98] K. L. Loftis, J. Price and P. J. Gillich. "Evolution of the Abbreviated Injury Scale: 1990-2015". In: *Traffic Inj Prev* 19.sup2 (2018), S109–S113. ISSN: 1538-957X (Electronic) 1538-9588 (Linking). DOI: 10.1080/15389588.2018.1512747. URL: https://www.ncbi.nlm.nih.gov/pubmed/30543458.

[99] TL Louw, G Kountouriotis, O Carsten and N Merat. "Driver Inattention During Vehicle Automation: How Does Driver Engagement Affect Resumption Of Control?" In: *White Rose Research Online* (2015).

[100] Marco Lützenberger. "A Driver's Mind". In: *Data Science and Simulation in Transportation Research*. Advances in Data Mining and Database Management. 2014. Chap. 10, pp. 182–205. ISBN: 9781466649200. DOI: 10.4018/978-1-4666-4920-0.ch010.

[101] Haneet Singh Mahajan, Thomas Bradley and Sudeep Pasricha. "Application of systems theoretic process analysis to a lane keeping assist system". In: *Reliability Engineering & System Safety* 167 (2017), pp. 177–183. ISSN: 09518320. DOI: 10.1016/j.ress.2017.05.037.

[102] Mail Online. *Warning to 223,000 Mini drivers after cars suffer sudden failure in power steering*. 2009. URL: https://www.dailymail.co.uk/news/article-1149132/Warning-223-000-Mini-drivers-cars-suffer-sudden-failure-power-steering.html (visited on 7th January 2022).

[103] Fulvio Mastrogiovanni, Antonio Sgorbissa and Renato Zaccaria. "A Distributed Architecture for Symbolic Data Fusion". In: (2007), pp. 2153–2158.

[104] Michael Matthews, David Bryant, Robert Webb and Joanne Harbluk. "Model for Situation Awareness and Driving: Application to Analysis and Research for Intelligent Transportation Systems". In: *Transportation*

Research Record: Journal of the Transportation Research Board 1779 (2001), pp. 26–32. ISSN: 0361-1981. DOI: 10.3141/1779-04.

[105] Natasha Merat, Hamish A. Jamson, Frank Lai and Oliver Carsten. "Human Factors of Highly Automated Driving: Results from the EASY and CityMobil Projects". In: (2014), pp. 113–125. ISSN: 2196-5544 2196-5552. DOI: 10.1007/978-3-319-05990-7_11.

[106] Natasha Merat and A Hamish Jamson. "How do drivers behave in a highly automated car". In: *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*. 2009, pp. 514–521.

[107] John A Michon. "A Critical View of Driver Behavior Models: What Do We Know, What Should We Do?" In: *Human Behavior and Traffic Safety* (1985), pp. 485–524. DOI: 10.1007/978-1-4613-2173-6-19.

[108] **Helen E. Monkhouse**, Ibrahim Habli and John McDermid. "An enhanced vehicle control model for assessing highly automated driving safety". In: *Reliability Engineering and System Safety* 202 (2020). ISSN: 09518320. DOI: 10.1016/j.ress.2020.107061.

[109] **Helen Monkhouse**, Ibrahim Habli and John McDermid. "The Notion of Controllability in an Autonomous Vehicle Context". In: *Critical Automotive applications: Robustness & Safety* (2015).

[110] **Helen Monkhouse**, Ibrahim Habli, John Mcdermid, Siddartha Khastgir and Gunwant Dhadyalla. "Why Functional Safety Experts Worry About Automotive Systems Having Increasing Autonomy". In: *Advanced & Trusted Computed*. IEEE, 2017.

[111] Xavier Mosquet, Michelle Andersen and Aakash Arora. "A roadmap to safer driving through advanced driver assistance systems". In: *Auto Tech Review* 5.7 (2016), pp. 20–25. ISSN: 2347-9434.

[112] Motor Industry Software Reliability Association. *Development Guidelines for Vehicle Based Software*. MIRA Limited, 1994.

[113] Motor Industry Software Reliability Association. *Guidelines for safety analysis of vehicle based programmable systems*. MIRA Limited, 2007. ISBN: 0952415674.

[114] Motor Industry Software Reliablity Association. *Use of Controllability for the Classification of Automotive Vehicle Hazards*. Report. Motor Industry Software Reliablity Association, 2007. URL: www.misra.org.uk.

[115] National Highway Traffic Safety Administration. *Automated driving systems: a vision for safety*. Tech. rep. DOT HS 812 442. U.S. Department of Transport, 2017. URL: https://www.nhtsa.gov/sites/nhtsa.gov/files/documents/13069a-ads2.0_090617_v9a_tag.pdf.

[116] National Institute of Health and Care Excellence. *Accessing Cost Effectiveness – Chapter 7 The Guidelines Manual*. URL: https://www.nice.org.uk/process/pmg6/chapter/assessing-cost-effectiveness.

[117] National Transport Safety Board. *Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian*. Journal Article. 2019. URL: https://www.ntsb.gov/investigations/accidentreports/reports/har1903.pdf.

[118] Alexandra Neukum, Eric Ufer, Jörn Paulig and HP Kruger. "Controllability of superposition steering system failures". In: *Steering tech* (2008).

[119] Christopher Nowakowski, Steven E. Shladover and Ching-Yao Chan. "Determining the Readiness of Automated Driving Systems for Public Operation: Development of Behavioral Competency Requirements". In: *Transportation Research Record: Journal of the Transportation Research Board* 2559.1 (2016), pp. 65–72. ISSN: 0361-1981 2169-4052. DOI: 10.3141/2559-08.

[120] Oxbotica. *Selenium explained*. 2021. URL: https://www.oxbotica.com/our-technology/ (visited on 10th February 2022).

[121] Oxford Language Dictionary. *"Communication"*. URL: https://languages.oup.com/google-dictionary-en/ (visited on 29th June 2022).

[122] Oxford Language Dictionary. *"Concept"*. URL: https://languages.oup.com/google-dictionary-en/.

[123] Oxford Language Dictionary. *"Geofence"*. URL: https://languages.oup.com/google-dictionary-en/ (visited on 16th January 2022).

[124] Oxford Language Dictionary. *"Model"*. URL: https://languages.oup.com/google-dictionary-en/ (visited on 2nd January 2023).

[125] Oxford Language Dictionary. *"Risk"*. URL: https://languages.oup.com/google-dictionary-en/ (visited on 8th April 2021).

[126] Oxford Language Dictionary. *"Uncertain"*. URL: https://languages.oup.com/google-dictionary-en/ (visited on 20th April 2021).

[127] Oxford Language Dictionary. *"Uncertainty"*. URL: https://languages.oup.com/google-dictionary-en/ (visited on 20th April 2021).

[128] Raja Parasuraman and Victor Riley. "Humans and automation: Use, misuse, disuse, abuse". In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 39.2 (1997), pp. 230–253. ISSN: 0018-7208.

[129] Polestar. *Connected Safety*. URL: https://www.polestar.com/uk/polestar-2/safety-features/driver-assistance-systems/ (visited on 1st April 2022).

[130] David Pumphrey. *The Principled Design of Computer System Safety Analyses*. Doctor of Philosophy. September 1999.

[131] J. Rasmussen. "Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models". In: *IEEE Transactions on Systems, Man, and Cybernetics* SMC-13.3 (1983), pp. 257–266. ISSN: 0018-9472. DOI: 10.1109/TSMC.1983.6313160.

[132] Roger Rivett. *Public Road Transport System and Vehicle Models*. Tech. rep. University of York, 2022.

[133] Bernd Rohrmann. "The risk notion: Epistemological and empirical considerations". In: *Integrative risk assessment* (1998), pp. 39–46.

[134] Paul M. Salmon, Neville A. Stanton, Guy H. Walker, Chris Baber, Daniel P. Jenkins, Richard McMaster and Mark S. Young. "What really is going on? Review of situation awareness models for individuals and teams". In: *Theoretical Issues in Ergonomics Science* 9.4 (2008), pp. 297–323. ISSN: 1463-922X 1464-536X. DOI: 10.1080/14639220701561775.

[135] Paul M. Salmon, Neville A. Stanton and Kristie Lee Young. "Situation awareness on the road: review, theoretical and methodological issues, and future directions". In: *Theoretical Issues in Ergonomics Science* 13.4 (2012), pp. 472–492. ISSN: 1463-922X. DOI: 10.1080/1463922X.2010.539289. URL: https://doi.org/10.1080/1463922X.2010.539289.

[136] Dario D Salvucci. "Modeling driver behavior in a cognitive architecture". In: *Human factors* 48.2 (2006), pp. 362–380. ISSN: 0018-7208.

[137] Robert Schubert and Marcus Obst. *The Role of Multisensor Environmental Perception for Automated Driving*. Ed. by Daniel Watzenig and Martin Horn. Automated Driving: safer and more efficient future. Springer, 2016. Chap. 7, pp. 161–182.

[138] Bobbie D. Seppelt and Victor W. Trent. "Potential Solutions to Human Factors Challenges in Road Vehicle Automation". In: *Road Vehicle Automation 3* (2016), pp. 131–148. ISSN: 2196-5544 2196-5552. DOI: 10.1007/978-3-319-40503-2_11.

[139] A. Shalom Hakkert and Victoria Gitelman. "Thinking about the history of road safety research: Past achievements and future challenges". In: *Transportation Research Part F: Traffic Psychology and Behaviour* 25 (2014), pp. 137–149. ISSN: 13698478. DOI: 10.1016/j.trf.2014.02.005.

[140] Paul Slovic. "Perception of risk". In: *Science* 236.4799 (1987), pp. 280–285. ISSN: 0036-8075.

[141] Society of Automotive Engineers. *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*. SAE J3016. Society of Automotive Engineers, 2016.

[142] Society of Automotive Engineers. *System Theoretic Process Analysis (STPA) Recommended Practices for Evaluations of Automotive Related Safety-Critical Systems*. Society of Automotive Engineers, 2022.

[143] Ankit Srivastava. *Sense-Plan-Act in Robotic Applications*. 2019. DOI: 10.13140/RG.2.2.21308.36481.

[144] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever and Ruslan Salakhutdinov. "Dropout: a simple way to prevent neural networks from overfitting". In: *The journal of machine learning research* 15.1 (2014), pp. 1929–1958. ISSN: 1532-4435.

[145] N. A. Stanton, P. M. Salmon, G. H. Walker, E. Salas and P. A. Hancock. "State-of-science: situation awareness in individuals, teams and systems". In: *Ergonomics* 60.4 (2017), pp. 449–466. ISSN: 1366-5847 (Electronic) 0014-0139 (Linking). DOI: 10.1080/00140139.2017.1278796. URL: https://www.ncbi.nlm.nih.gov/pubmed/28051356.

[146] N. A. Stanton et al. "Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology". In: *Ergonomics* 49.12-13 (2006), pp. 1288–311. ISSN: 0014-0139 (Print) 0014-0139 (Linking). DOI: 10.1080/00140130600612762. URL: https://www.ncbi.nlm.nih.gov/pubmed/17008257.

[147] Neville Stanton. "Models and methods for collision analysis". In: *Safety Science* (2019).

[148] Neville A Stanton. "Why do self-driving cars crash? Reactions to automated driving". In: *Day 2 Key Note Speech*. SAFEComp. September 2021.

[149] Neville A. Stanton and Philip Marsden. "From fly-by-wire to drive-by-wire: Safety implications of automation in vehicles". In: *Safety Science* 24.1 (1996), pp. 35–49. ISSN: 0925-7535. DOI: http://dx.doi.org/10.1016/S0925-7535(96)00067-7. URL: http://www.sciencedirect.com/science/article/pii/S0925753596000677.

[150] Chauncey Starr. "Social Benefit versus Technological Risk". In: *Science* 165.3899 (1969), pp. 1232–1238. ISSN: 00368075, 10959203. URL: http://www.jstor.org/stable/1727970.

[151] Alan Stevens, Corinne Brusque and Joseph Krems. *Driver Adaptation to Information and Assistance Systems*. The Institution of Engineering and Technology, 2013, 382 pp. ISBN: 978-1-84919-639-0.

[152] Jack Stilgoe. "How can we know a self-driving car is safe?" In: *Ethics and Information Technology* (2021). ISSN: 1388-1957 1572-8439. DOI: 10.1007/s10676-021-09602-1.

[153] D. L. Strayer and D. L. Fisher. "SPIDER: A Framework for Understanding Driver Distraction". In: *Human Factors* 58.1 (2016), pp. 5–12. ISSN: 0018-7208. DOI: 10.1177/0018720815619074. URL: %3CGo%20to%20ISI%3E://WOS:000370709100001.

[154] L. J. Suchman. "What is Human-Machine Interaction". In: *Cognition, Computing, and cooperation*. Ed. by S. P. Robertson, Z. Wayne and J. B. Black. Vol. 2. Intellect Books, 1990. Chap. 3.

[155] Mark A. Sujan, Ibrahim Habli, Tim P. Kelly, Astrid Gühnemann, Simone Pozzi and Christopher W. Johnson. "How can health care organisations make and justify decisions about risk reduction? Lessons from a cross-industry review and a health care stakeholder consensus development process". In: *Reliability Engineering & System Safety* 161 (2017), pp. 1–11. ISSN: 0951-8320. DOI: http://dx.doi.org/10.1016/j.ress.2017.01.001. URL: //www.sciencedirect.com/science/article/pii/S0951832017300017.

[156] Zehang Sun, George Bebis and Ronald Miller. "On-road vehicle detection: A review". In: *IEEE transactions on pattern analysis and machine intelligence* 28.5 (2006), pp. 694–711. ISSN: 0162-8828.

[157] T E Solution. *Wind Tunnel Test for Buildings by TE Solution*. URL: http://www.tesolution.com/windtunnelbuildings.html (visited on 2nd January 2023).

[158] Tesla. *Autopilot and Full Self-Driving Capability*. 2021. URL: https://www.tesla.com/en_GB/support/autopilot-and-full-self-driving-capability (visited on 12th February 2021).

[159] *Tesla Model 3: Paris' largest taxi firm suspends cars after fatal crash*. December 2021. URL: https://www.bbc.co.uk/news/world-europe-59647069 (visited on 21st January 2022).

[160] The American Association of State Highway and Transportation Officials. *Highway Safety Manual*. 2010. URL: http://www.highwaysafetymanual.org/Pages/default.aspx (visited on 2nd January 2022).

[161] The Assurance Case Working Group. *Goal Structuring Notation Community Standard*. Tech. rep. Safety Critical Systems Club, 2018.

[162] The International Organization for Standardization. *ISO 26262: Road Vehicles–Functional safety*. Vol. ISO 26262. 2018.

[163] The International Organization for Standardization. *ISO PAS 21448: Road vehicles – Safety of the intended functionality*. Publically Available Specification. 2018.

[164] The International Organization for Standardization. *ISO/SAE 21434: 2021 Road Vehicles – Cybersecurity engineering*. Standard. 2021.

[165] The International Organization for Standardization. *ISO 21448: Road vehicles – Safety of the intended functionality*. Draft international standard. 2022.

[166] The Royal Academy of Engineering. *Societal Aspects of Risk*. Report. The Royal Academy of Engineering (London), 2003.

[167] The Tesla Team. *Upgrading Autopilot: Seeing the World in Radar*. 2016. URL: www.tesla.com/en_GB/blog/upgrading-autopilot-seeing-world-radar (visited on 13th November 2016).

[168] The World Cafe. *World Cafe Method*. Tech. rep. The World Cafe, 2017. URL: http://www.theworldcafe.com/key-concepts-resources/world-cafe-method/ (visited on 15th January 2017).

[169] TheoryTest.org. *Pedestrian Crossings*. URL: https://theorytest.org.uk/pedestrian-crossings/ (visited on 3rd April 2022).

[170] Paul B. Thompson. "THE Philosophical Foundations Of Risk". In: *The Southern Journal of Philosophy* 24.2 (1986), pp. 273–286. ISSN: 0038-4283. DOI: https://doi.org/10.1111/j.2041-6962.1986.tb01566.x.

[171] Paul B. Thompson. "Risk Objectivism and Risk Subjectivism: When Are Risks Real". In: *Risk: Issues in Health and Safety 1* No. 1 (1990), pp. 3–22.

[172] Paul B. Thompson. "Ethics and Risk Communication". In: *Science Communication* 34.5 (2012), pp. 618–641. ISSN: 1075-5470 1552-8545. DOI: 10.1177/1075547012459177.

[173] Eric Thorn, Shawn Kimmel and Michelle Chaka. *A Framework for Automated Driving System Testable Cases and Scenarios*. National Highway Traffic Safety Administration, 2018.

[174] Yvonne Toft, Geoffrey Dell, KK Klockner and Allison Hutton. "Models of causation: safety". In: *Safety Institute of Australia, Tullamarine, Victoria* (2012).

[175] Traffic accident research at the TU Dresden. *German In-depth Accident Study*. 2022. URL: https://www.vufo.de/gidas-pcm/?L=1 (visited on 23rd January 2022).

[176] David Undercoffler. *Crashing The Self-Driving Party Of Tesla & Co*. December 2021. URL: https://www.forbes.com/sites/davidundercoffler/2021/12/30/crashing-the-self-driving-party-of-tesla-et-al/?sh=27027fbf4a82 (visited on 10th February 2022).

[177] UNECE. *WP.29 – Introduction*. URL: https://unece.org/wp29-introduction (visited on 7th January 2022).

[178] United Nations Economic and Social Council. *Vienna convention on road traffic*. United Nations Economic and Social Council, November 1968.

[179] United Nations Economic and Social Council. *Resolution R.E.6 on the administrative and technical provisions required for carrying out the technical inspections according to the technical prescriptions specified in Rules annexed to the 1997 Agreement*. United Nations Economic and Social Council, April 2017. URL: https://unece.org/fileadmin/DAM/trans/main/wp29/wp29resolutions/ECE-TRANS-WP29-1132e.pdf.

[180] United Nations Economic and Social Council. *Uniform provisions concerning the approval of vehicles with regard to Automated Lane Keeping Systems*. UNECE 157 Revision 3. United Nations Economic and Social Council, March 2021. URL: https://unece.org/sites/default/files/2021-03/R157e.pdf.

[181] US Department of Defense. "System safety program requirements". In: *MIL-STD-882c, US Department of Defense, USA* (1993).

[182] U.S. Food and Drugs Administration. *Economic Impact Analyses of FDA Regulations*. URL: https://www.fda.gov/about-fda/reports/economic-impact-analyses-fda-regulations (visited on 29th May 2021).

[183] T. W. Victor, E. Tivesten, P. Gustavsson, J. Johansson, F. Sangberg and M. Ljung Aust. "Automation Expectation Mismatch: Incorrect Prediction Despite Eyes on Threat and Hands on Wheel". In: *Human Factors* 60.8 (2018), pp. 1095–1116. DOI: 10.1177/0018720818788164.

[184] Nicole A. Vincent. "A Structured Taxonomy of Responsibility Concepts". In: *Moral Responsibility*. Library of Ethics and Applied Philosophy. 2011. Chap. Chapter 2, pp. 15–35. ISBN: 978-94-007-1877-7. DOI: 10.1007/978-94-007-1878-4_2.

[185] G. H. Walker, H. Gibson, N. A. Stanton, C. Baber, P. Salmon and D. Green. "Event Analysis of Systemic Teamwork (EAST): a novel integration of ergonomics methods to analyse C4i activity". In: *Ergonomics* 49.12-13 (2006), pp. 1345–69. ISSN: 0014-0139 (Print) 0014-0139 (Linking). DOI: 10.1080/00140130600612846. URL: https://www.ncbi.nlm.nih.gov/pubmed/17008260.

[186] Guy H Walker, Neville A Stanton and Paul M Salmon. *Human Factors in Automotive Engineering and Technology*. CRC Press, 2017. ISBN: 978-1-138-74725-8.

[187] James C. Walliser, Ewart J. de Visser, Eva Wiese and Tyler H. Shaw. "Team Structure and Team Building Improve Human–Machine Teaming With Autonomous Agents". In: *Journal of Cognitive Engineering and Decision Making* 13.4 (2019), pp. 258–278. ISSN: 1555-3434. DOI: 10.1177/1555343419867563.

[188] Wenshuo Wang, Xiaoxiang Na, Dongpu Cao, Jianwei Gong, Junqiang Xi, Yang Xing and Fei-Yue Wang. "Decision-making in driver-automation shared control: A review and perspectives". In: *IEEE/CAA Journal of Automatica Sinica* (2020), pp. 1–19. ISSN: 2329-9266 2329-9274. DOI: 10.1109/jas.2020.1003294.

[189] Waymo. *Waymo Safety Report*. Journal Article. 2018. URL: https://waymo.com/safety/ (visited on 10th February 2022).

[190] *What does it take? A compilation of Tesla autopilot failures and predictable abuses*. October 2019. URL: https://www.youtube.com/watch?v=ZqP14aGjdj8 (visited on 21st January 2022).

[191] Wikipedia. *Concept — Wikipedia, The Free Encyclopaedia*. URL: https://en.wikipedia.org/wiki/Concept (visited on 2nd January 2023).

[192] Wikipedia. *Mental Model — Wikipedia, The Free Encyclopaedia*. URL: https://en.wikipedia.org/wiki/Mental_model (visited on 24th January 2023).

[193] Wikipedia. *National Traffic and Motor Vehicle Safety Act — Wikipedia, The Free Encyclopaedia*. URL: https://en.wikipedia.org/wiki/National_Traffic_and_Motor_Vehicle_Safety_Act (visited on 7th January 2022).

[194] Wikipedia. *OODA Loop — Wikipedia, The Free Encyclopaedia*. URL: https://en.wikipedia.org/wiki/OODA_loop (visited on 9th January 2022).

[195] Wikipedia. *The Theory of Mind — Wikipedia, The Free Encyclopaedia*. URL: https://en.wikipedia.org/wiki/Theory_of_mind (visited on 14th February 2023).

[196] Wikipedia. *Chernobyl Disaster — Wikipedia, The Free Encyclopaedia*. 2016. URL: https://en.wikipedia.org/wiki/Chernobyl_disaster (visited on 29th December 2016).

[197] Wikipedia. *Haddon Matrix — Wikipedia, The Free Encyclopaedia*. April 2021. URL: https://en.wikipedia.org/wiki/Haddon_Matrix (visited on 1st January 2022).

[198] Wikipedia. *Vehicle Regulation — Wikipedia, The Free Encyclopaedia*. 2021. URL: https://en.wikipedia.org/wiki/Vehicle_regulation (visited on 7th January 2022).

[199] Tom Wilson. *Sensing Technologies for the Autonomous Vehicle*. 2016. URL: https://www.slideshare.net/embeddedvision/sensing-technologies-for-the-autonomous-vehicle-a-presentation-from-nxp-semiconductors?

qid=a716eeb3-c336-449f-8726-9011cea09825&v=&b=&from_search=3 (visited on 10th February 2022).

[200] Matthew Wood et al. *Safety First for Automated Driving*. Aptiv Services US, 2019.

[201] World Health Organization. *Road Traffic Injuries*. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries (visited on 8th January 2022).

[202] World Road Association. *Road Safety Manual*. 2019. URL: https://roadsafety.piarc.org/en (visited on 2nd January 2022).

[203] Y. Xing, C. Huang and C. Lv. "Driver-Automation Collaboration for Automated Vehicles: A Review of Human-Centered Shared Control". In: *2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1964–1971. DOI: {10.1109/IV47402.2020.9304755}.

[204] Robert K Yin. *Case study research : design and methods*. eng. 5th ed. Los Angeles ; London : SAGE, 2014. ISBN: 9781452242569.

[205] M. S. Young, N. A. Stanton and D. Harris. "Driving automation: Learning from aviation about design philosophies". In: *International Journal of Vehicle Design* 45.3 (2007), pp. 323–338. DOI: 10.1504/IJVD.2007.014908.

[206] Jens O. Zinn. *Social Theories of Risk and Uncertainty : An Introduction (Chapter 7)*. Hoboken, UK: John Wiley & Sons, Incorporated, 2008. ISBN: 9781444301496. URL: http://ebookcentral.proquest.com/lib/york-ebooks/detail.action?docID=416512.

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

`https://bitbucket.org/amiede/classicthesis/`

Final Version as of 16th March 2023 (`classicthesis` Version 2).